SEARCH FOR NEW PHYSICS IN EVENTS WITH COLLIMATED PHOTONS AND GLUONS

By

SAVVAS KYRIACOU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

written under the direction of

Yuri Gershtein

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2019

ABSTRACT OF THE DISSERTATION

Search for new physics in events with collimated photons and gluons

By SAVVAS KYRIACOU

Dissertation Director:

Yuri Gershtein

This thesis presents the design and results of a search for physics beyond the standard model in events containing jets with substructure resulting from new particles decaying into a photon and two gluons. Jet substructure techniques are adapted to develop a new approach for photon identification in a dense hadronic environment. The analyzed proton-proton collision data were collected by the CMS experiment at center-of-mass energy of 13 TeV in 2016 and correspond to an integrated luminosity of 35.9 inverse femtobarns. No statistically significant excess is observed in data. The first cross section limits on new physics processes resulting in such events are set and interpreted in the context of a Stealth Supersymmetry model.

Acknowledgments

This has been a long trip and a transformative experience. I would like to thank everyone that shared some moments in this path with me.

Especially, I would like to thank my supervisor Yuri Gershtein for caring, providing guidance and inspiration and being patient with me the 7 years, this undertaking has lasted. I believe it's imperative to have a motivated and passionate supervisor and Prof. Gershtein has been both, helping me overcome all the obstacles in the way.

A big thanks to Alejandro Gomez for being a good friend and a partner in crime all this time. We have not found SUSY yet Ale, but it is definitely not dead. Elliot Hughes thank you for your interesting and intelligent perspectives.

I would also like to thank my parents, Athena and Andreas and my siblings for everything.

Alexandra,Spencer and Vilmi thank you for being there.

Dedication

To the beauty guiding us

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Despite the success of the standard model of particle physics, there are numerous indications, such as the cosmological observations of dark matter, the measured value of the Higgs boson mass [2], as well as theoretical and aesthetic considerations such as naturalness, that suggest the existence of new physics at the TeV energy scale. The LHC has not uncovered signs of new physics thus far. Searches for new phenomena have been focusing on signatures of large energy imbalance in the detector, that could indicate escaping undetected new states. Another approach followed up to now, was to select events based on how many well reconstructed and isolated lepton and photons they contain, to suppress the SM background processes. This type of searches may not be sensitive to new phenomena that could be hidden in the background SM processes with large cross sections. The most common such processes at the LHC, quark/gluon scattering, result in multijet events. These processes are listed in a hierarchical order in fig. 1.1. This leaves a big space for the new physics to live in and yet to be undetected.

There is a plethora of well-motivated theoretical scenarios that predict the appearance of new physics in the background dominated regions, with events not having either large energy imbalance signatures nor isolated leptons and photons. These signal events would appear similar to multijet events at the LHC. The models that predict such events include hidden valley models [49, 48] and a number of supersymmetric (SUSY) models, such as R parity violating SUSY [9] and stealth SUSY [27, 26, 25]. The analysis presented in this thesis, focuses on a Stealth Supersymmetric model that predicts events with jets composed of photons and gluons and large hadronic activity in the signal events. We focus on uncovering

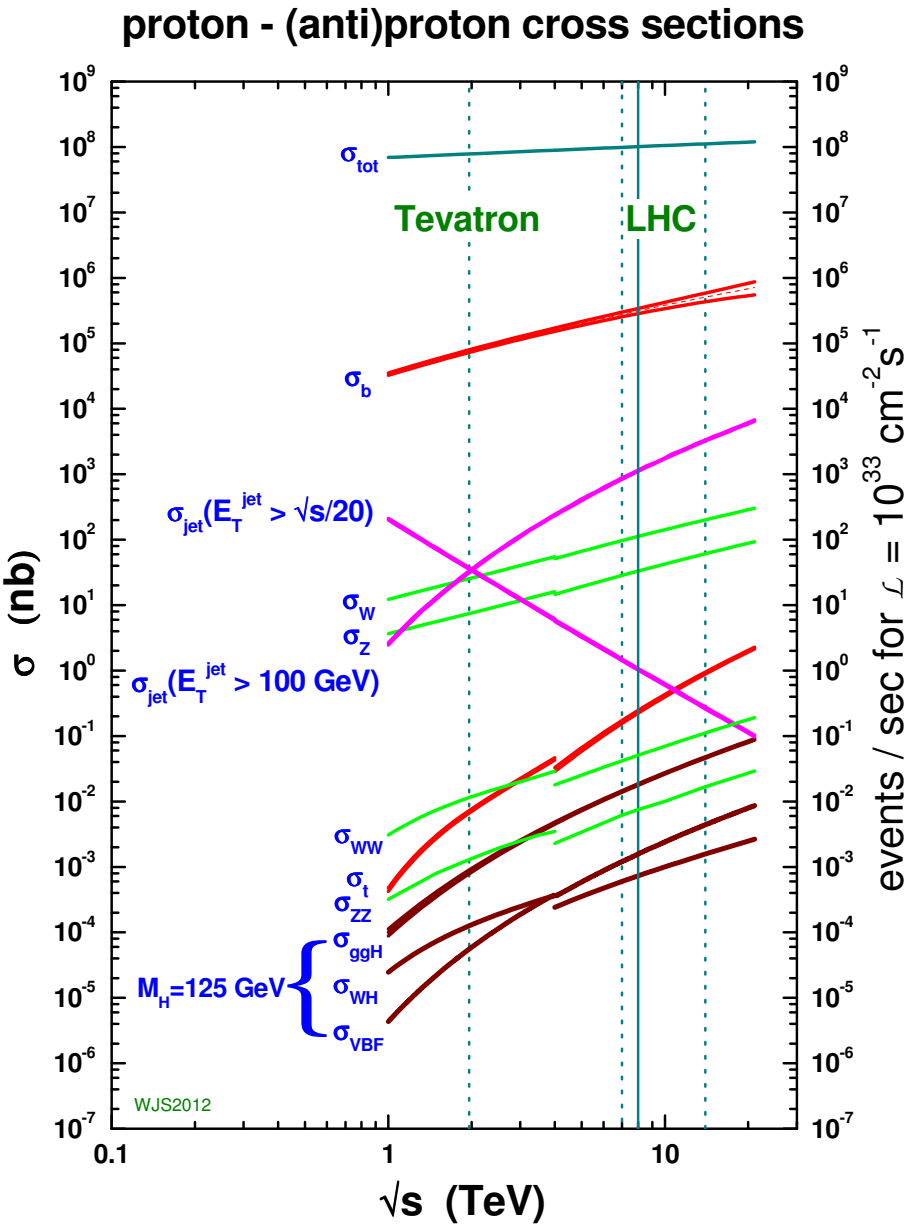

Figure 1.1: Cross sections for different SM processes vs center-of-mass energy for proton -proton collisions. The figure is taken from [47]

events that contain pairs of unucual jets composed of photons and gluons and this way suppress the standard model process and discover any signs of new physics.

In this chapter, the fundamental fields and their interactions are outlined. A brief outline of the Standard model is presented in the first section. Following that, extensions to the standard model such as supersymmetry are described. We focus on Stealth Supersymmetry and the details of the benchmark model we used for this search. Finally motivations for the search we performed are outlined and the main characteristics of the analysis are presented.

1.1 The Standard Model of Particle Physics

The standard model (SM) of particle physics is the most studied, well tested and successful physical theory in existence. It describes the fundamental interactions of quantum fields and is the closer thing existing to a Theory of Everything. It obtained it's current form in the late 70's after the discovery of the first quarks and it's last piece ,the Higgs bosons was found in 2012 (when I was beginning my PhD). Quantum field theory is the child of quantum mechanics and special relativity. Or differently a story of constraining theories with specific symmetries that lead to the appearance of new interactions and mechanisms.

Elementary particles can be bosons or fermions. Bosons have integer spin and abide to the Bose-Einstein statistics while fermions have half integer spins and obey the Pauli exclusion principle. SM fermions come in three generations of leptons and quarks. The interactions of leptons and quarks are outlined in fig. 1.3. Leptons are fundamental particles that are not color-charged and thus do not interact via the strong interaction, while they have electroweak charge and thus interact under weak and electromagnetic interactions (with the exception of neutrinos that interact only through the weak force). Quarks, except for the electroweak charge they carry, they are also charged under the strong interaction. This charge has been poetically named color and thus the term Quantum Chromodynamics (QCD) - the dynamics of color - describing the strong interaction. The prediction and subsequent discovery of the third generation of quarks was one of the big successes of the

standard model.

The first generation of leptons, is made from the electron flavor with the two types being the electron it self and the other the neutrino. The second is made of the muon flavor with the corresponding muon lepton and muon neutrino, while the third is the tau lepton and neutrino. In fig. 1.2 the leptons are presented with their masses listed. The first generation of quarks, is what makes protons and neutrons for the most part. The up and down quarks. The second generation included the charm and strange quarks and the third is made of the top and bottom quarks with the top, being the heaviest of all and the last that has been discovered. The top quark unlike the rest of the quarks, is so heavy that it doesn't form bound states (mesons and baryons) and has a minute lifetime, decaying to the rest of the quarks.

Bosons are the mediators of the fundamental interactions. The photon is the mediator of the electromagnetic interaction, it is massless and thus the interaction has an infinite range and is propagated with the speed of light. Based on the fact that gravity also has an infinite range, affecting the large scale dynamics of the universe and it was measured to be propagated with the speed of light, gravitons - which have not been discovered, are assumed to have zero mass. Unlike gravity and electromagnetism, the weak interaction is mediated by the W and Z bosons. These have large masses (80 GeV and 90 GeV respectively) and based on the uncertainty principle, one can deduce the range of this force, which is confined well within the nucleon radius. Gluons are the mediators of the strong interaction and are massless. Unlike the electroweak interaction, quantum chromodynamics (QCD) is a non-Abelian theory, that states that the mediators of the theory interact with each other at the tree level. This has profound consequences. One would expect the range of the strong interaction based on the mediator mass to be infinite but because of confinement, the range of the strong interaction is limited to radius of a nucleus. This also results in quarks and colored charged particle, to be always confined inside mesons or other hadrons where the total color charge is canceled and never be bare.

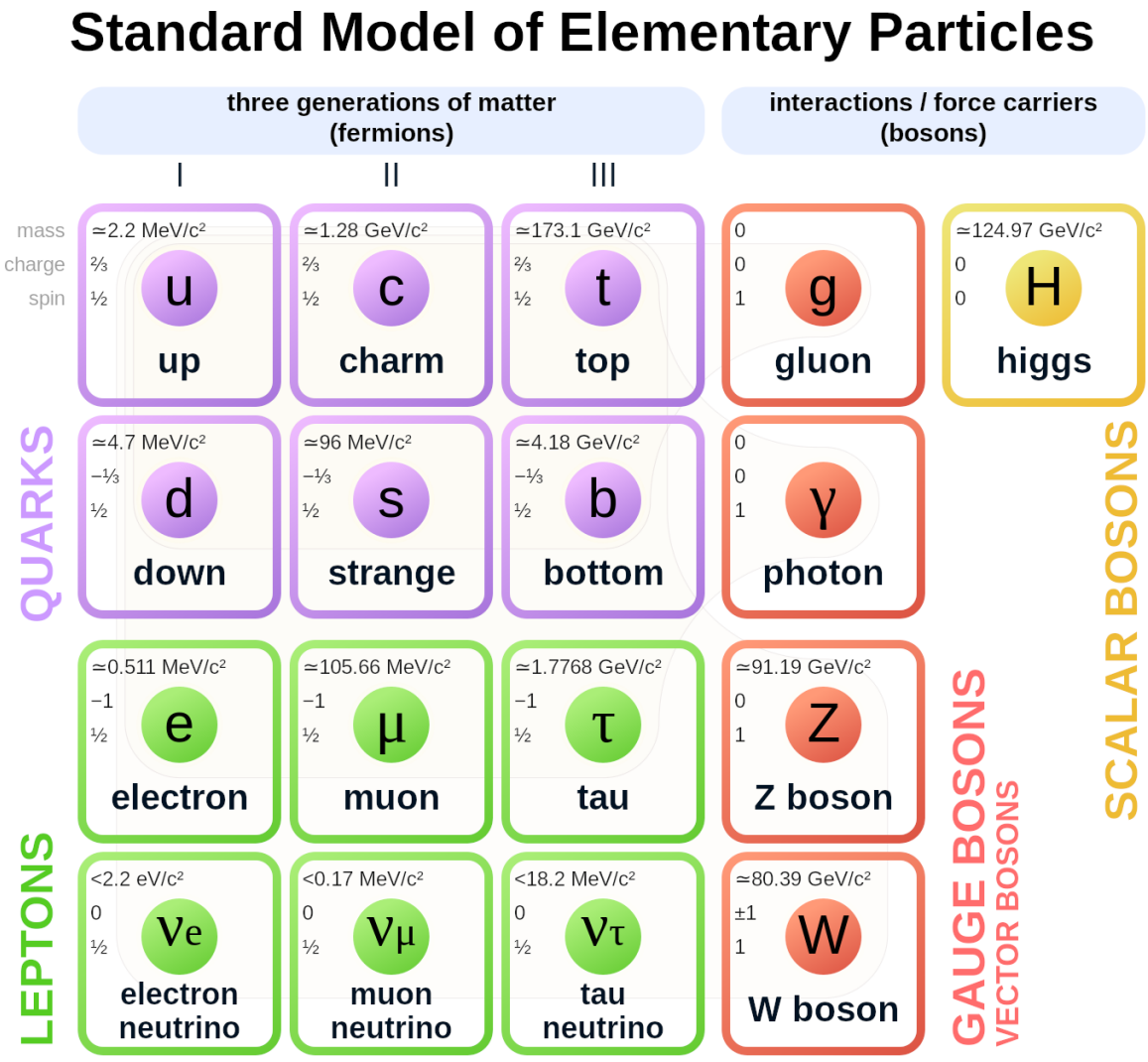

Figure 1.2: The Fundamental particles of the Standard Model. The masses, charge and spin of each is noted and the particles are grouped in their respective species generations and types. The image is taken from [22].

The last elementary bosonic field predicted by the SM to be discovered is the Higgs field. This discovery is the greatest achievement up to date by the LHC experiments. The Higgs boson is a massive spin 0 particle. Through the Higgs mechanism, the electorweak symmetry is broken and the electromagnetic and weak interactions appear as distinct forces at low energy scales. This is achieved with the W and Z bosons acquiring masses. The Higgs mechanism, is one case of spontaneous symmetry breaking, where event though a theory has a symmetry in it's Lagrangian, it's ground state doesn't have this symmetry property.

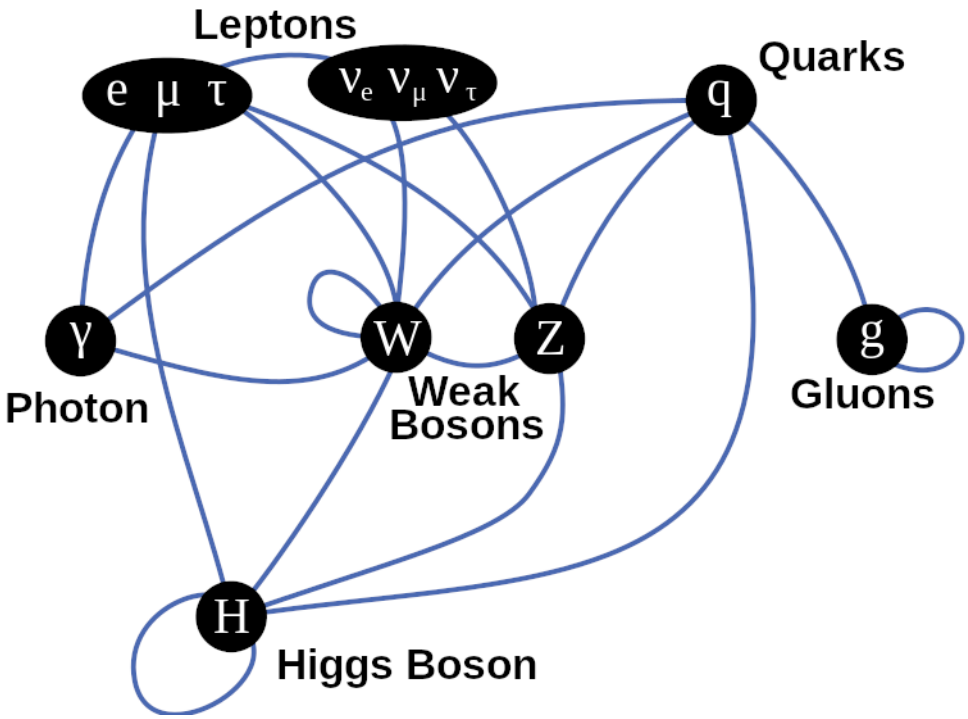

Figure 1.3: The interactions between the elementary bosons and fermions. A particle connected to the gluons (g) interacts with the strong force. A particle connected to Wand Z bosons interacts with the weak force and a particle connected to the photon(γ) interacts with the electromagnetic force. The image is taken from [21]

Attempts to cast gravity in the same framework of a quantum field theory has proven to be a problem, that just like calculating pi, resulted in an explosion of new techniques, tools and ideas. The SM thus, does not describe gravity. When one attempts to write a quantum field theory of gravity, infinities appear that can not be renormalized. This problem though, lead to the creation of string theory and other beautiful ideas that attempt and

some achieve to quantize gravity and unify it with the rest of the forces that still remain to be experimentally tested.

Beyond the failure of the SM's framework to describe gravity and unify it with the rest of the forces, there are a number of indications that this is not the last word in the story of finding a theory that describes the universe. Among those, there are two indications coming from cosmological observations. The first is the existence of an unknown state of matter that permeates galaxies and causes stars to orbit faster than one would expect based on their distance from the galactic center. This matter because of the fact that it does not interact electromagnetically, simply put, it is not emitting electromagnetic radiation that we observe unlike baryonic matter in the universe, was named dark matter. Several independent lines of evidence point to the existence of this new state of matter (or states), including beyond the rotation curves from galaxies, evidence in the cosmic microwave background power spectrum, measurement of galaxy masses from gravitational lensing, the bullet cluster and indications from galactic simulations that show that with out dark matter, the gravity of the baryonic matter is too weak to form galaxies. The second cosmological indication for physics beyond the standard model, is the observation of the accelerating expansion of the universe that point to the existence of the so called dark energy. Multiple models attempt to describe this as a quantum field that got switched on and contributed to the expansion rate, against gravity after the Big Bang.

Other direct indication for physics beyond the SM are neutrino oscillations and CP violation not accounted by the SM. Tensions also exist between measurements of CP violation in b meson decays and the new observation(2019) of CP violation in the charm sector with the SM that need to be resolved and indicate that new forces could exist. The observed matter anti-matter asymmetric in the universe is not accounted by the CP violation predicted by the SM.

The SM has 19 free parameters. These parameters, are not fixed by any mechanism and appear to be arbitrary. One could consider this a sign of an incomplete theory. The Higgs

mass, one of these free parameters, the value of was measured by the LHC experiments in 2012 to be close to 125 GeV. The Higgs mass is sensitive to quantum corrections that an expected value of considering these corrections would be immensely bigger than the observed one. On the other hand, the Higgs mass within the SM is expected to be between below 1 TeV for the SM to be consistent [1], []. This issue along with the energy scale of the electroweak symmetry breaking compared to the Planck scale fall under the general term, the hierarchy problem.

1.2 Beyond the Standard Model

There are two ways one can perceive the hierarchy between the electroweak scale and the Planck scale -or why the mass of the Higgs is so light. The one path assumes that this is a result of fine tuning, that is that the parameters of the standard model are such that this scale appears as such just purely because of luck and coincidence. The other perspective calls for the existence of a mechanism that creates this hierarchy and fixes the Higgs mass to the measured value. The other possibility is that the energy scale difference between the electroweak and Planck scale is not as stark and the existence of extra dimensions places the Planck scale to the TeV range. The mechanisms that can provide an explanation to this problem are usually theories that extend the standard model by introducing new interactions, states and phenomena. Some of these theories attempt to solve multiple of the open problems of the standard model.

Supersymmetry [39] is one of the most attractive and also well studied family of BSM theories. It predicts the existence of a symmetry between the two families of fundamental particles, the bosons and fermions. It predicts that a mirror set of the SM fundamental particles exist with similar properties but with fermions having bosonic mirror partners and the opposite. These superpartners of the SM particles are expected to be heavier in mass, since supersymmetry is a broken symmetry if it exists, at the TeV energy range, pushing the scale where this symmetry is restored in higher energies. SUSY provides a possible solution

to the hierarchy problem, with the introduction of the superparticles. The superparticles are predicted to interfere destructively and cancel out the Higgs mass terms coming from SM particles like that top quark that would otherwise blow up the value. Because of this cancellation, the Higgs mass is protected and the fine tuning of the standard model parameters to fix the Higgs mass to the observed value is not necessary. Another appealing feature of SUSY is that it predicts the unification of the fundamental forces at high energies leading to a Grand Unified Theory (GUT) as a step in the direction of also integrating gravity into the Theory of Everything. In fig. 1.4 the running coupling constants of the fundamental interactions predicted by SUSY are presented. It might be naive though to assume that the only interactions existing in nature are the ones we have observed up to now and future results could indicate otherwise, which GUT's should accommodate for.

Supersymmetry is a broken symmetry and different models have to assume how it is broken. This results in a multitude of approaches and a very diverse phenomenology. R-parity is defined as particle quantum number, that takes a +1 value for SM particles and a -1 value for supersymmetric particles. The decays of the the SUSY particles to lighter ones and SM particles, create decay chains that end up in the Lightest SUSY partner(LSP). In the R-Parity conserving SUSY models (RPC), the LSP is stable and is a possible dark matter candidate. These particles are predicted to escape the detector creating signatures of missing transverse momentum. Compressed Spectra SUSY and Stealth SUSY are among the RPC SUSY models, that through different mechanisms predict events without the smoking gun signature of missing ET. In PRV SUSY, the LSP may decay to SM particles. In some cases, only decays to quarks are allowed leading to events with no missing transverse momentum.

In the stealth SUSY framework [27, 26, 25], a hidden sector of particles is postulated, that doesn't couple to the SUSY breaking mechanism and thus the symmetry still holds there. The particles and their superpateners in this sector are expected to be almost mass degenerate and decay to each other through the emission of gravitinos (assumed to be the

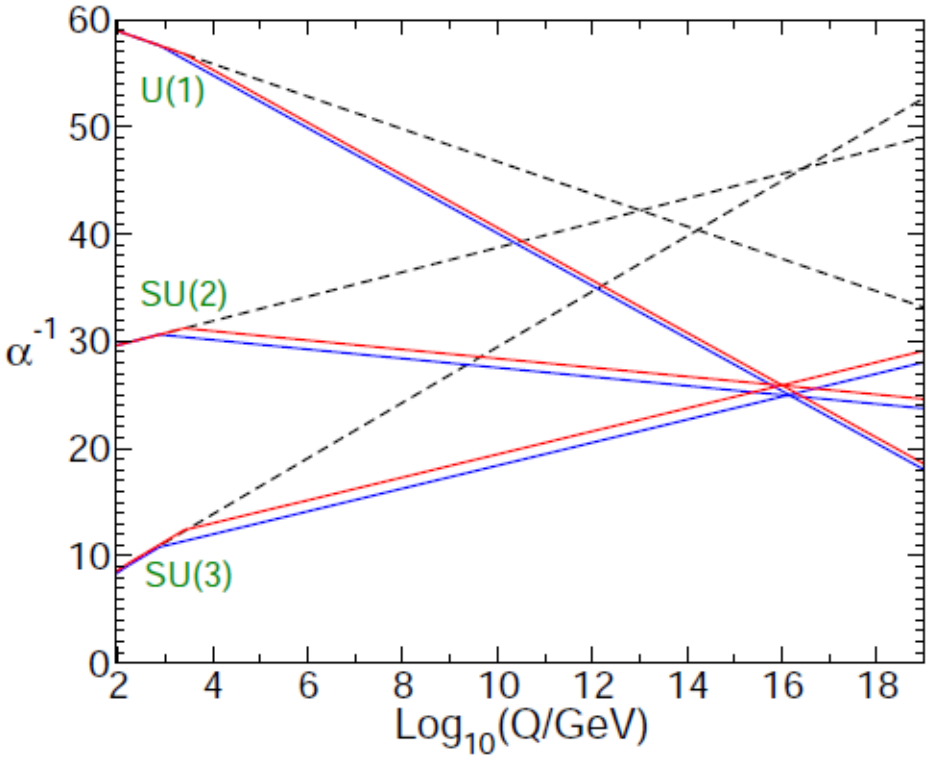

Figure 1.4: The running coupling constants of the fundamental interactions as predicted by a SUSY model known as MSSM(colored lines) and the SM(dashed line). U(1) denotes the electromagnetic interaction, SU(2) the weak interaction and SU(3) the strong interaction. It is predicted that the forces unify at large energies. The x axis presents the energy scale and the y axis the inverse of the coupling constant. The diagram is taken from [39]

LSP) that would carry the supersymmetric R-Parity charge and be very soft compared to the hidden sector particle momenta, thus eliminating missing transverse energy signatures. In the search presented in this thesis, we assume a minimal Stealth SUSY model, in which the hidden sector includes only a pair of superpartners. A mass gap of 10 GeV between the two hidden sector superparteners ensures that the energy imbalance in the event will be limited. The specifics of the model we use are presented in the next section.

1.3 The search

We focus on the pair production and decay of gluinos according to the diagram presented in fig. 1.5. The diagram shows how one gluino produced would decay according to Stealth SUSY. Two of these decay chains are expected to be observed per event in the detector. Gluinos ($\widetilde{\mathrm{g}}$), the gluon superpartners, are expected to be created with large cross sections at the LHC and decay to neutralinos $\widetilde{\chi}_1^0$ and a quark-antiquark pair. Stealth SUSY assumes gauginos (either neutralinos or charginos), which decay to a $\tilde{S}$ and a photon(γ), to be the portal to the hidden-sector. The $\tilde{S}$ is expected to decay to an S and a gravitino ($\widetilde{\mathrm{G}}$) with the subsequent decay of the S to a pair of gluons. Because of the mass degeneracy of the hidden-sector pair, the $\widetilde{\mathrm{G}}$ is expected to be produced with low momentum and thus the event to be characterized with low missing transverse momentum. The cross sections assumed in this analysis for gluino pair production, are presented here, while we assume that the branching fraction of gluinos to the decay chain we study is one.

Previous searches at CMS for stealth SUSY [34, 15] required two isolated photons. The isolation requirement reduces the sensitivity to cases where a mass hierarchy exists between the electroweak gauginos, in this case the $\widetilde{\chi}_1^0$, and the colored superparticle ($\widetilde{\mathrm{g}}$). If this large mass hierarchy is present, the $\widetilde{\chi}_1^0$ is expected to be produced with a large Lorentz boost and its decay products to be collimated, resulting in non-isolated photons. This mass hierarchy seems to be suggested by the current LHC results, since the exclusion limits for colored gluinos and squarks, extend well above the TeV mark, while exclusion limits for gauginos,

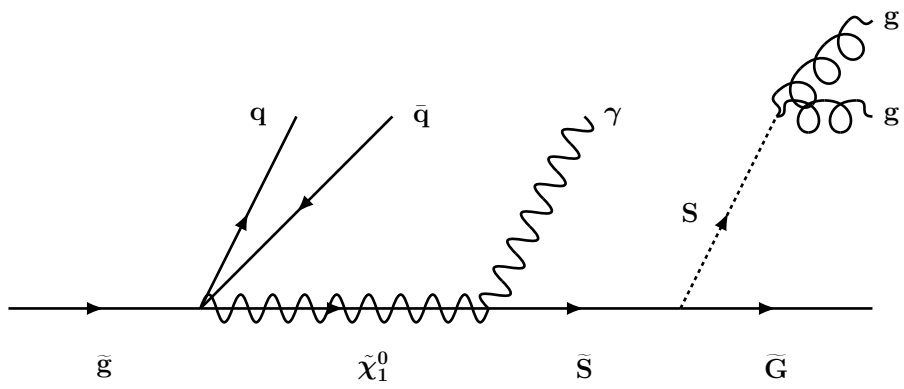

Figure 1.5: Single leg diagram describing the decay chain of a gluino ($\widetilde{\mathrm{g}}$). The decay products of the $\widetilde{\chi}_1^0$ are all expected to be merged in a single jet. That is a jet composed of a photon (γ) and two gluons originating from the S decay. Each signal event is expected to have two such decay chains coming from the two produced gluinos. This analysis uses this process as the targeted signal signature for discovery.

are not as strong.

We follow a complementary strategy, to previous searches, by identifying jets composed of one photon from the $\widetilde{\chi}_1^0$ decay and a pair of gluons from the S decay, which we refer to as *photon jets*. It is possible to identify photon jets by utilizing a combination of existing and novel jet substructure tools, that we present in this thesis. Within the minimal stealth SUSY model we consider, superparticle production at the LHC is predicted to appear as events that have two photon jets associated with a large hadronic activity in the event. The distribution of the total hadronic activity of events containing photon jets is used to perform hypothesis testing and search for evidence of new physics, possibly hidden in the SM multijet background.

Chapter 2

The LHC and CMS

The LHC [24] with its experiments, CMS, ATLAS, ALICE and LHCb is an astonishing machine, one of the biggest scientific undertakings humanity has ever attempted and achieved. A rich physics program is ongoing at the LHC, ranging from generic BSM searches and SM measurements to studies of the quark gluon plasma and b decays. With the end of the Run2 of the LHC, the four experiments have completed and are still producing measurements using the proton-proton collisions as also the Heavy Ion collisions provided by the LHC. The highlight of the physics searches at the LHC was the discovery of the the last missing piece of the Standard Model, the Higgs boson.

The accelerator is located outside the city of Geneva in Switzerland, crossing the Swiss-French border and it is housed in the tunnels of the old LEP collider [40]. A complex system of linear accelerators and boosters are used to increase the energy of proton beams in stages to 450GeV before they are injected into the LHC. Initially protons are accelerated by a linear collider - the LINAC followed by the booster , the proton synchrotron (PS) and the super proton synchrotron(SPS). The protons are split in two beams while passing through this complex, circulating in opposite directions. A schematic of the accelerator complex is presented in fig. 2.1. The LHC has a circumference of 27km stationed with 1232 superconducting dipole magnets creating a magnetic field of 8T needed to steer the charged particles around the ring. Another 392 quadrupole magnets are used to focus the beams while a number of radiofrequency (RF) cavities inject energy in the particle beams, with the potential to increase the final beam energy to 7TeV. The collider uses proton beams and not beams of opposite charge particles and this necessitates the magnetic fields acting

on the two beams to have opposite polarities. This is achieved with the dipole magnets intricate design presented in fig. 2.2 . The proton beams are split in bunches with a 25ns time gap between them. Each bunch contains about 10^11 protons. The two proton beams cross each other at four interaction points where the collider experiments are located to measure and record the particle collisions.

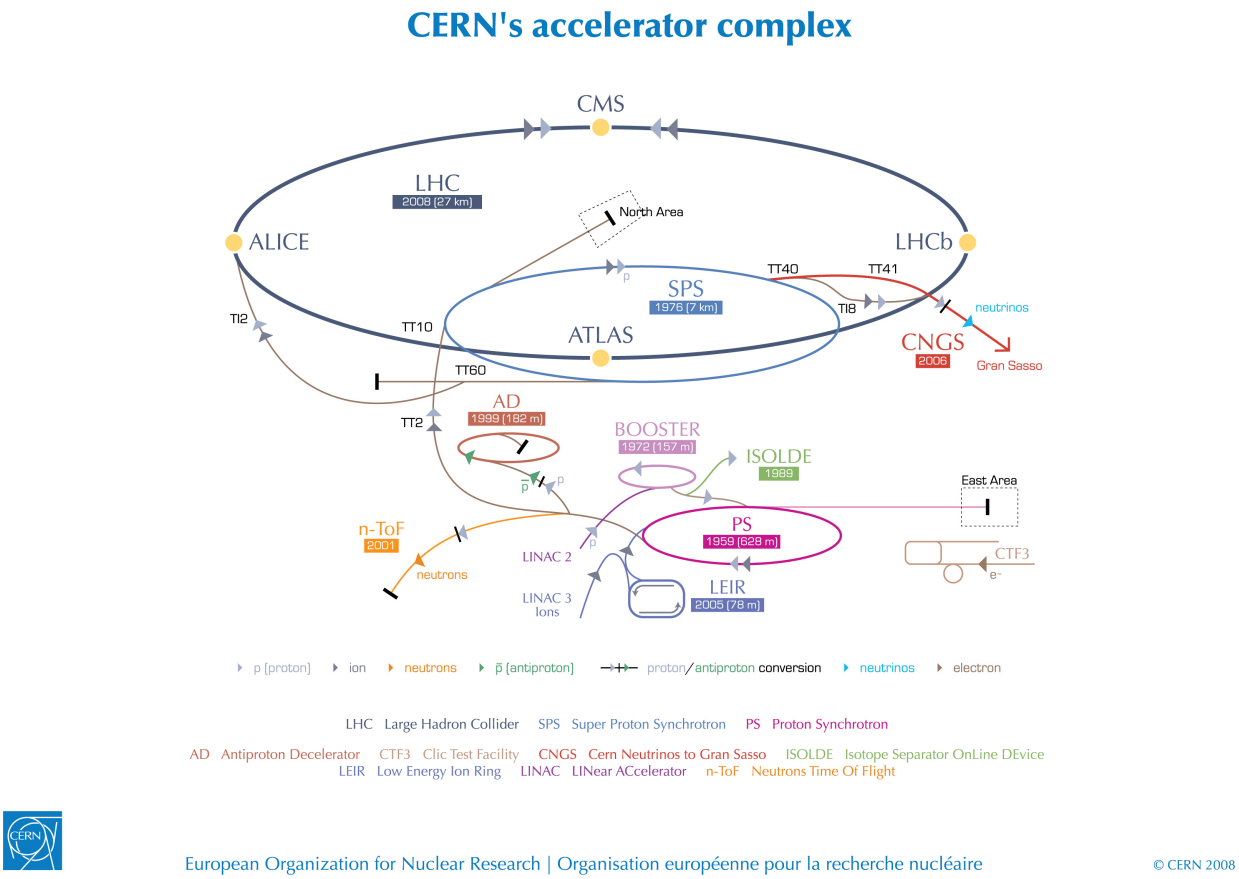

Figure 2.1: The CERN accelerator complex, leading to the Large Hadron Collider. Each individual accelerator, raises the beam energies to be injected in the next machine. The image is taken from [38]

Having a proton-proton (pp) beams instead of electron-positron beams allows one to increase the energy reach of the accelerator machine since less energy is radiated away by the more massive protons in comparison to the light electrons (such as in LEP), as the particles are accelerated. A higher center of mass energy, is thus achievable, that allows for heavier states to be created by the collisions. The LHC has already performed runs with 7 TeV

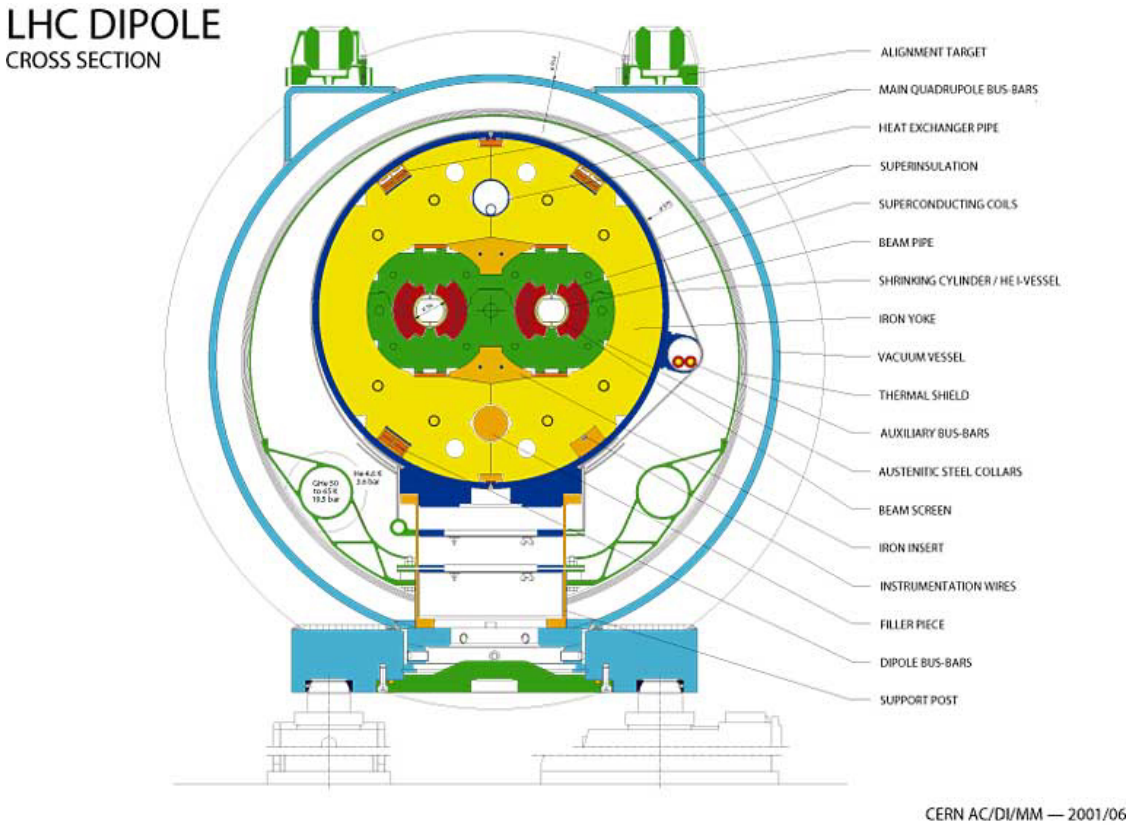

Figure 2.2: A cross section of an LHC dipole magnet. In the center two beam pipes exist, one for each of the two proton beams circulating around the collider in opposite directions. The image is take from [23]

and 8 TeV center-of-mass energy while the Run 2 which the analysis presented in this thesis focuses on, was performed with a 13 TeV center-of-mass energy. This brings the LHC to the energy frontier of collider physics, with no other existing collider having this energy reach. A disadvantage of pp colliders compared to electron-positron machines, is that because of the structure of the proton and the parton distribution functions, interacting partons from the two protons, scan a range of energies that they interact with. Thus, observing the Higgs decays to invisible particles and measuring a branching fraction on this process, which can be used as a test for the Higgs coupling to possible dark matter candidates, can be better performed at a lepton collider.

Unlike the Tevatron collider [52] that used proton - anti-proton beams, the LHC is a symmetric collider. Proton beams are easier to make than anti-proton beams, where the anti-protons must first be produced through colliding proton beams on a target, store the anti-protons in a ring and then accelerated them. This means a higher number of particles per beam for the LHC, reaching record instantaneous luminosity levels. As a result the LHC is now at the luminosity frontier. The instantaneous luminosity in a circular collider is defined as

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^\star} F(\theta) \tag{2.1}$$

where N is the number of protons per beam, n_b is the number of bunches, f_{rev} is the rotation frequency of the bunches in the collider while γ_r is the relativistic γ factor. The denominator describes the geometric cross-section of the beams. This includes the hourglass effect factor ($\beta^\star$), while the $\tilde{F}$ describes the distribution of partons in the proton. The LHC has managed to collect an order of magnitude more data during the LHC Run2 than the Tevatron during its entire life time. This provides the opportunity to study extremely rare particle physics processes that were not accessible before.

The p-p initial state favors color interactions. Beyond the rich QCD color interactions, the fact that the proton is a composite object and the high instantaneous luminosity at

the LHC create ”dirty” events with a multiplicity of final states and multiple soft and hard scatters that can inundate the processes of interest. This multiple scattering instances per bunch crossing, caused by the high instantaneous luminosity are what we particle physicists call as Pileup and create a background that needs to be subtracted from events of interests. Pileup mitigation has been and will be one of the biggest challenges the LHC experiments have faced and will face with the High Luminosity upgrade of the collider. Nonetheless, the LHC is now, the most energetic and highest luminosity collider humanity ever had to study fundamental physics at the smallest scales.

2.1 The Compact Muon Solenoid Apparatus

The CMS experiment is located at the P5 interaction point of the LHC in France and it is a complex assemblage of sub systems and detectors with the purpose of recording the collisions provided by the LHC [11]. It has a cylindrical structure with different detecting systems layered over others in layers, around the interaction point. The detector has a superconducting solenoid that provides a 3.8T magnetic field, used to measure the properties of the particles. The experiment has the ability to detect a number of final state particles, such as leptons, photons and jets utilizing a multitude of detector technologies. This renders CMS, with the ATLAS experiment, one of the most versatile collider experiments at the LHC with a broad physics reach. A general schematic of CMS is presented in fig. 2.3.

2.2 Kinematic variables

Before describing the detecting subsystems of CMS it is imperative to describe the kinematic variables and coordinate system collider experiments use.

The direction of a vector in cylindrical collider experiment, is measured with two variables. The first is the angle ϕ (azimuth), that is measured as the vector direction in the vertical plane, perpendicular to the beam. The second variable is a measure of the polar angle (θ) with respect to the beam axis. The variable that measures the polar angle is named

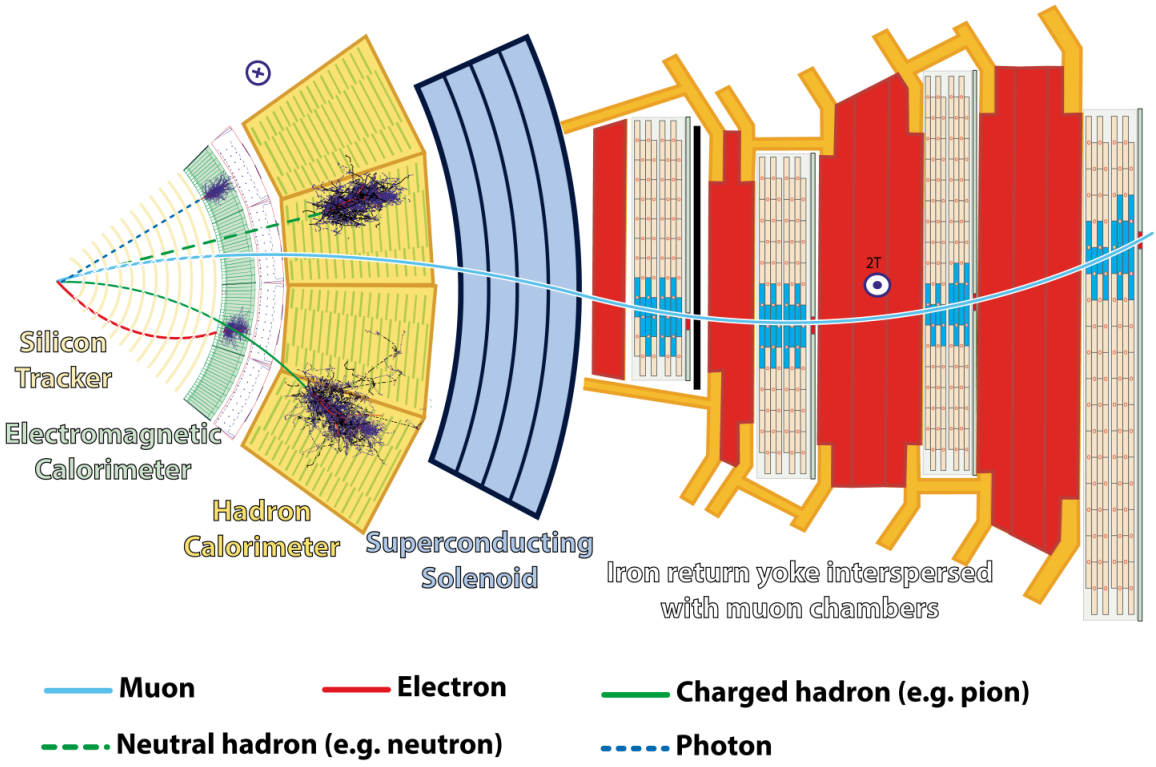

Figure 2.3: A cross section of the CMS detector where all the different subsystems are indicated. It is also demonstrated how different types of particles interact with detector. The diagram is taken from [10]

pseudorapidity and it is denoted with the greek letter η. Pseudorapidity (η) is defined as:

$$\eta = -ln(\frac{\theta}{2}) \tag{2.2}$$

Pseudorapidity ranges from $\pm$ inf for vectors pointing along the beam lines. An anglular distance used is the the so called ΔR variable. This is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.3}$$

where, $\Delta\eta$ is the difference in the pseudorapidity and $\Delta\phi$ the difference in ϕ between two points in the $\eta - \phi$ space.

Another useful variable used by collider experiments, is the that of transevrse momentum defined as the projection of the momentum of a particle on the perpendicular plane to the beam line. This quantity is denoted as p_{T}.

2.2.1 The Inner Tracking Systems

Tracking at CMS [20] is performed by a combination of detectors, with the two inner most of them being the Pixel and the Silicon Strip detectors. Both systems rely on Silicon sensors, in which incoming charged particles create electron-hole pairs which can be then detected. The magnet creates a 3.8T homogeneous field in the total volume of the tracker, in which particles bend and momentum measurements are feasible, while trapping low momentum particles in the tracker from inundating the rest of the detector systems. The task of the tracking system is to provide excellent resolution and efficiency for the 3D track and secondary vertex reconstruction. In the LHC environment with the instantaneous luminosity of $10^{34}\ cm^{-2}s^{-1}$ the challenge is also to avoid high occupancies and radiation degradation effects. In the current collider detector paradigm design, where tracking systems are placed between the interaction area and the calorimeters, a low tracking material budget is necessary to reduce photon conversion, multiple scattering and bremsstrahlung from the

incoming particles. At the same time the systems must have enough structural support for the detecting elements, cooling systems and readout electronics. The η coverage of the Pixel and SST detectors is between -2.5 to 2.5.

The Pixel Detector

The pixel detector consists of 3 barrel layers and 2 disks (3 disks after the 2017 upgrade) on each side with about 65 million readout channels. The first barrel layer is placed at a radius of 4.4 cm, at 7.3 cm the second and the third at 10.2 cm. The disks are placed at 34.5 and 46.5 cm along the z axis. The detector employs silicon sensors each with dimensions of 100 μm width and 150 μm length while the thickness of the sensors is 250 μm. The readout electronics are places on the back of the sensitive material connected with indium bump bonds. Charge sharing between pixel elements allow for a finer resolution than the size of the elements them selfs at the level of 15 μm.

Silicon Strip Tracker Detector

The Silicon Strip tracker is divided in 4 sub sections. The Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker Outer Barrel (TOB) and the Tracker End Caps (TECs). The TIB consists of four barrel layers starting at a radius of 25.5 cm and the last layer at 49.8 cm complemented by 3 disks of the TID system in each side extending radially from 20 to 50 cm away from the beam line. The outer barrel has six layers extending to 1.16m and extends in z between $\pm$ 118 cm. TEC has 9 disks extending radially from 22 cm to 1.14m. In total, there are 11.4 million strip elements in the four subsystems with less granularity than the pixel detector. The detector resolution is η depended and given by :

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = (15 p_{\mathrm{T}} 0.5)\%(\,\mathrm{TeV}) for |\eta| < 1.4 \tag{2.4}$$

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = (60 p_{\mathrm{T}} 0.5)\%(\ \mathrm{TeV}) for |\eta| > 1.4 \tag{2.5}$$

2.2.2 Calorimetry

Calorimetry is one of the oldest technologies used to detect particles by forcing incoming particles in a material to shower and thus deposit energy in the detector. The material captures a fraction or the total energy of the developed shower and re-emits it as light based on the phenomenon of scintillation. The emitted light is then captured by photosensitive devises and converted to electric signals that are processed. Calorimeters can also utilize Cerenkov radiation to measure the particle properties. Such detectors are usually calibrated by using mono-energetic particle beams and mapping the electric signal output to the incoming particles energy. Homogenous calorimeters are detectors in which a single volume is used to stop the incoming particle and measure the totality of its energy, while sampling calorimeters are detectors composed of layers of sensitive and non sensitive material and measure a fraction of the total energy of the particle's shower. Calorimeters used in collider experiments are usually structurally split into two systems, an electromagnetic (ECAL) one, that is placed closer to the interaction point and are used to measure the energy of photons and electrons and a hadronic one (HCAL), that is placed after the electromagnetic calorimeter and is used to contain strongly interacting particles, such as pions, neutrons and other hadrons. The ECAL and most of the HCAL in CMS are inside the bore of the superconducting magnet in the barrel region ($|\eta| < 1.47$). The constrain on the size of the detecting elements, to fit in the magnet bore, sets stringent limitations to the material selection and design of the detector modules.

Electromagnetic Calorimeter

The CMS electromagnetic calorimeter [17], a homogeneous calorimeter, was designed with the purpose of having an excellent energy resolution, to allow for the discovery of the

Higgs boson by measuring energy of the final state photons. The detecting material used is lead tungsten ($PbW0_4$) which has a small Moliere radius and an extremely short radiation length (X_0). It is also transparent to the scintillation light generated in it. Scintillation light is produced in the $PbW0_4$ crystals as a response to the electromagnetic showers by photons, electrons and other energy depositions by particles passing through or stopping in the detector. The structure of the EM calorimeter is divided into two pseudorapidity sections each with its own layout. The Barrel region that covers the region up to $|\eta| < 1.47$ and the Endcap section that extends up to $|\eta| < 3$. An overall schematic of the ECAL is presented in fig. 2.4.

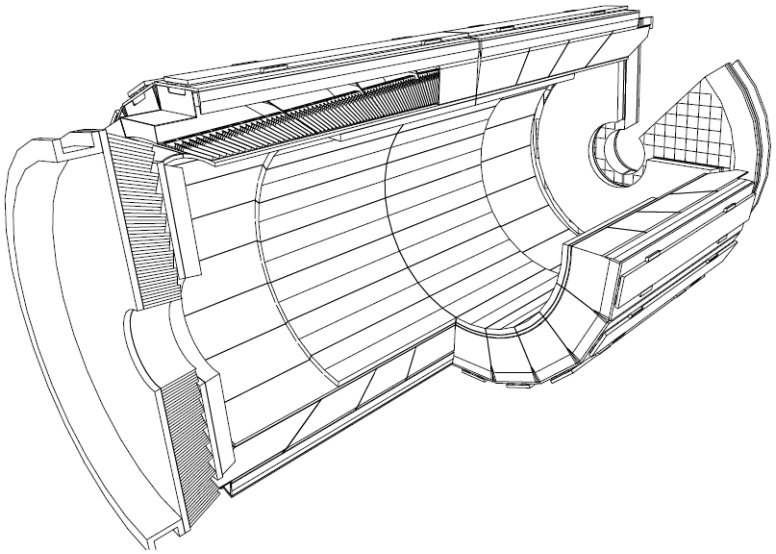

Figure 2.4: A schematic of the electromagnetic calorimeter and how the crystals are grouped to form supermodules and rings to form the barrel. One of the Ds is also visible with the pre-shower detector attached to it, that comprise the end cap electromagnetic calorimeter. The image taken from [17].

The Barrel section, is made of 61200 lead tungsten crystals and has an inner radius of 1.3 m. Each crystal covers a $\Delta\eta - \Delta\phi$ area of 0.0175x0.0175. This area corresponds to 22 by 22 mm and covers the Moliere radius of lead-tungsten that is 21.9 mm. Crystals have a length of 23 cm that is about 26 χ_0 in the central region. The crystals are directed to point to the detector center, with a small offset of 3° to ensure hermeticity. Avalanche

photodiodes (APD), convert the scintillation light from the crystals into electric signals that are fed to calorimeter read out system. The selection of the APD technology for the light measurement is based on the insensitivity of the APD elements to high magnetic fields as the one found in the central region of the detector.

A total number of 7324 crystals arranged into two D shapes compose each end cap part of the EM calorimeter. The area of each crystal in the end cap is approximately 28.6x28.6 mm arranged in an x-y grid. The crystals are arranged in 5x5 groups called super crystals. The length of each endcap crystals is shorter by 3 cm compared to the Barrel crystals but the presence of the pre-shower detector (ES) right in-front of the end cap EM calorimeter increases the total radiation lengths to 26 as in the Barrel. The readout of scintillation light detection of the endcap crystals is done using vacuum phototriodes. The pre-shower consists of two detecting silicon layers and lead layers in between them with the purpose of discerning neutral pions, photons and electrons reaching the detector.

The transparency of the crystal to scintillation light has been found to deteriorate based on the radiation exposure with a highly non predictable behavior. To mitigate this effect and limit the resolution uncertainties because of the transparency loss, a laser system is used to inject light into the crystals. The response to the injected light, is used to measure the crystal transparency . Generally many effects affect the performance of the ECAL. The resolution of energy measurements by the ECAL is given by the formula :

$$(\frac{\sigma}{E})^2 = (\frac{2.8\%}{\sqrt{E}})^2 + (\frac{0.12}{E})^2 + (1\%)^2 \tag{2.6}$$

where the first term corresponds to stochastic effects, such as fluctuations in the lateral containment of the EM showers, fluctuation in the number of photo-electrons released in the APDs for given energies and fluctuations in the energy deposited in the pre-shower to the pre-shower reading. The sceond term in the equation refers to noise originating from the read-out electronics and pile-up effects and the last term is the constant term. The

constant term described containment effects, non uniform light yield and inter-calibration errors.

Hadronic Calorimeter

The hadronic calorimeter [18] is designed to detect strongly interacting particles, that initiate hadronic showers deep in the detecting material. Shower leaks from the ECAL are also contained by the HCAL and the energy deposition profiles in these two detectors is used to discriminate electromagnetic objects from charged or neutral hadrons. The detector is split in 4 sub-detectors, each covering complementary and different η and radial regions. A sampling technology with absorber layers and sensitive scintillator layers is used. The HCAL has the highest η coverage and it is therefore essential to measuring the Missing energy in the event.

The inner rapidity region, $|\eta| < 1.4$ and inside the magnet bore, is covered by the HCAL barrel detector (HB). This detector consists of brass absorber layers with plastic scintillating layers in-between them. The selection of brass is based on the non magnetic nature of the material and its interaction length. The layers of absorber and sensitive material are organized into towers, with all the sensitive layers being read as one and cover a $\Delta\eta - \Delta\phi$ area of 0.087x0.087. There are 17 plastic scintillator layers with brass absorber in between them and steel layers for the most inner and outer absorbers for structural support. Towers are organized into 18 wedges that cover 20 degrees in ϕ. The sensitive layers of a single tower are all read as one with relative weighting to account for thickness differences. The sensitive layers have a pitch of a few mm and the absorber material thickness varies between 4 to 7 cm. A wave length shifting fiber is used in the detector that is then spliced with a clean fiber the transmits the scintillation light to the light detectors to be transformed into electronic signals. The HB is complemented by the Hadronic Outer (HO) calorimeter which is placed outside the magnet and the detector $\eta - \phi$ granularity is kept the same to the HB detector. The HO purpose is to measure high energy hadronic showers that leak

outside the HB or are initiated outside it. It covers up to $|\eta| < 1.3$. The HB and HO provide about 16 interaction lengths combined in the central region.

Following the same general architecture as the ECAL detector, the HB is complemented by endcap hadronic stations known as HE. This detector covers a pseudorapidity region between 1.3 and 3. Brass and plastic scintillators are also used for the absorber and detecting material as in the HB. The tower setup also follows the HB with the towers areas starting at $\Delta\eta - \Delta\phi$ 0.087x0.087 and increasing to about 0.17x0.17 .

In the most forward region and at 11.2 m in z from the detector center the Hadronic Forward calorimeter HF extends the rapidity coverage to $|\eta| < 5$. Steel is used as the absorber material and quartz fibers as the sensitive material, with the purpose of measuring Cherenkov radiation initiated in the fibers by flying particles. The choice of these materials is based on the high particle fluence in the high forward region from beam remnants and other radiation. This detector is used also as an instantaneous luminosity estimator complementing the pixel luminosity telescope.

In the HCAL system, the stochastic term dominates the energy resolution with the constant term following. The resolution is described by the formula :

$$(\frac{\sigma}{E})^2 = (\frac{90\%}{\sqrt{E}})^2 + (4.5\%)^2 for |\eta| < 3.0 \tag{2.7}$$

$$(\frac{\sigma}{E})^2 = (\frac{172\%}{\sqrt{E}})^2 + (9.0\%)^2 for |\eta| > 3.0 \tag{2.8}$$

2.2.3 Muon Systems

The Muon systems in CMS [19] have a threefold task. To identify muon candidates originating from the interaction point, act as a muon triggering system and measure the momentum and sign of candidate muons. This is achieved by combining information from the muon systems and other detecting systems such as the tracker, described above. Since muons are

minimum ionizing particles, they traverse the detectors embeded in the magnetic flux return yoke and can be thus identified as tracks extending outside the tracker and calorimeters. The muon detectors are thus positioned outside the HO, in the magnetic field return yoke.

The technology selection and designs for the muon detection systems, was based on particle rates in different detector areas, rapidity coverage, magnetic field intensity and time responses while keeping the number of channels low. Three different gaseous detecting technologies are used. The mechanism behind the function of gaseous detectors is that muons or other charged particles, traversing the gas will ionize the gas and in the presence of an electric field, an avalanche of charge with be formed which can be detected to obtain spatial and timing measurements for the particle track. Drift Tube detectors (DTs) are used for covering the barrel region, Cathode Strip Chambers (CSCs) for the End Cap region and Resistive plate chambers (RPCs) compliment the other two technologies both in the Barrel and End Cap.

Drift Tubes are essentially tubes filled with gas and a wire traversing them in the middle, with a voltage applied to the wire and tube to guide the avalanche electrons to the wire and positive ions to the tube. The size of the tube and voltage applied sets the detector reaction time, since bigger tubes, mean larger drift times for the charge to reach the wire. This modular technology, with decoupled detecting elements (the tubes) can provide bunch crossing identification and a 100 micron hit resolution per tube. The tubes are arranged in layers of parallel wires and 4 DTs layers form a muon station.

Cathode Stripe Chambers are a detectors composed of an anode wire plane and cathode strip plane oriented perpendicularity to the wire plane. These two planes are embedded in a gas mixture and they provide two spatial measurements for each traversing particle. CSCs can provide precise position and timing measurements in the high particle flux in the End Cap and magnetic field intensity.

RPCs consist of two resistive material plates that have a small gap between them and the outer surface of each is painted with a conducting graphite paint making them into electrodes

with high voltage applied. Readout is performed by aluminum strips. This detector when run in avalanche mode, provides a timing resolution at the order of nanoseconds with the drawback being low signal pulses that the front end electronics have to amplify. Beyond the excellent timing resolution, this detector can be highly segmented providing sufficient spatial measurement of the incoming tracks making them ideal for triggering purposes.

In the Barrel region $|\eta| < 1.2$, four DT stations are arranged within and out of the magnetic field return yoke, two in the outside surfaces and two inside the yoke. The stations follow the structure of the iron return yoke and thus there are 5 rings of them placed longitudinally, along the bean line. The stations are further segmented into chambers in these rings, with 60 chambers in the inner 3 stations and 70 in the outer station in total. Each chamber has 12 layers of DTs and these are grouped into three SuperLayers - four DTs layers form a super layer, stacked with half a tube height difference. The two first superlayers are used to measure the ϕ coordinate of the muon track while the third is used for the measurement of the z parameter of the track. Beyond the DT chambers, RPC layers are mounted with the DTs for triggering purposes. There are six layers of RPC chambers in total in the barrel , with 2 of them in the inner two stations and one for each of the outer stations.

The End Cap is comprised our stations of Cathode Strip chambers. These are composed of concentric rings of trapezoidal chambers centered along the beam line. There are 3 such rings for the first muon endcap station while two rings for each of the other stations. The outer radially rings for stations 2nd the 4th station, have 36 trapezoidal chambers covering an angle of 10 deg while the inner ring has 18 such chambers covering each 20 deg. To ensure hermeticity, all the chambers overlap in ϕ. The maximum η coverage the CSCs provide is 2.4. RPCs are also present in the End Cap with four layers present, one layer for each of the CSC stations. The rapidity coverage of the RPC layers in the End Cap reaches up to 2.1.

2.2.4 Triggering System

At the designed instantaneous luminosity of the LHC of $10^{34} cm^{-2} s^{-1}$, the proton-proton interaction frequency is huge exceeding hundreds of MHz. Most of these interactions are inelastic QCD scattering events, of well studied processes. Ideally CMS could record all the bunch crosses and then search for interesting events in the collected dataset. The problem lies in the fact that this is a huge rate for recording, storing and studying everything the collider produces. Interesting processes related to new physics are expected to happen with a much lower rate than the large SM cross section processes. The purpose of the triggering system is to reject the ordinary and record the extraordinary processes that show up in the detector. This mean that the system has to understand what is present in the event and based on pre-determined criteria decide if a process is worth recording for further analysis or it should be rejected.

This task, is achieved with a two tiered system [37]. The first tier filters events to a rate of 100kHz and it is based on custom electronic boards that provide event decisions within 4μs after the bunch-crossing. The second tier reduces the rate of recorded events to 1KHz and it is based on commercial computers. The first system is known as the Level-1 Triggering system (L1), while the second is the High Level Trigger (HLT). It is necessary to ensure that a signal process that is searched for, will pass the pre-determined selections, the triggering system applies to record an event. Limitations from the triggering system, are reflected to the physics program of an experiment and a clever design can allow for discovery.

Chapter 3

Object Reconstruction

Every recorded collision by the detector, is a set of electronic readings from the different subdetecting units described in the previous chapter. Particles traversing the detector deposit energy in the different subsystems generating these electronic signals. To make sense of what the detector recorded, and identify what final state particles have been created in each event, these electronic signals have to be combined by algorithms into objects. This processes is referred to as reconstruction. In this chapter the algorithms for the reconstruction of objects relevant to this analysis, are presented. Flowing that, general kinematic variables that are useful for analyzing collision data are defined. Reconstruction has a two fold task, on the one hand perform particle identification and thus tell if a particle is a muon or electron and on the other hand measure the kinematic and other properties for the observed particles.

3.1 Objects Reconstruction

To reconstruct a particular kind of final state particle from the detector readings, one has to address which subdetectors this particle has interacted with. That is, an electron is expected to leave hits in the tracking material of the detector and then a cluster of energy in the ECAL. There are a number of collections of reconstructed objects from the detector, that can refer to the same kind of particles, ie an electron or a photon, but differ in the specifics algorithm that was used to reconstruct them and how and what information was combined to do so.

In Run2 of the LHC, the Particle Flow (PF) algorithm [43], a framework that reconstructs final state objects using information from the whole experiment, helped to improve the performance of the detector with respect to previous algorithms and techniques used. The idea behind particle flow is to reconstruct every final state particle created as a unique such particle candidate based on the information from the tracking systems, the calorimeters and the muon systems. The particle candidates reconstructed by the PF algorithm, are named PF candidates. The resulting particle flow candidates, can then be clustered to form complex objects like jets. The utilization of the pixel and outer tracker detector with their phenomenal resolution and reconstruction capabilities, combined with the excellent ECAL granularity, is what boosted the overall reconstruction performance.

3.1.1 Tracks

Track fitting is performed in CMS with a Kalman Filter procedure [16]. A track will be first seeded, that is initiated by a set of close hits in the tracking material that are consistent with each other either with momentum or direction criteria. For example three sequential hits in the pixel layers in small angular distance from each other can be identified as a track seed. Track seeds from the pixel detector, are named pixel seeds and this variable is used further on in this analysis. The seed then is extrapolated in the forward and backwards directions to identify hits that could belong to a track consistent with the track seed. The candidate hits in other tracker layers will be checked for consistency with the already selected set of hits. Iteratively the track parameters will be updated with each hit addition using a Bayesian approach. Because of the magnetic field that the tracker is embedded in, charge particles, that leave hits in the tracker, have bended tracks, and the momentum of the charged particle can be measured.

3.1.2 Electrons

Electrons are expected to leave a hits in the tracker and then shower and get absorbed by the CMS calorimeters. Electron tracks, are reconstructed with hits from both the pixel and outer tracker, fit to a 5 parameter helix [32]. The track then is associated with a cluster of energy in the electromagnetic calorimeter of CMS. As the electron propagates through the tracking material, it can also radiate photons because of its interaction with the material, a process known as bremsstrahlung. These photons then can either propagate to the calorimeter or convert to electron positron pairs that then reach the calorimeter where they are absorbed. This makes electron track reconstruction a more complex task, since the helix track fitter has to include corrections for energy losses of this kind. A dedicated track fitter known as Gaussian Sum Filter (GSF) takes care of the radiation effects.

3.1.3 Photons

Photons unlike electrons, are not charged and are not expected to leave hits in the tracking material. Thus they are expected to propagate through the tracker and then deposit their energy in the ECAL. Crystals with energy depositions are then clustered together and form superclusters that measure the total energy of the photon.

Because of the tracking material, photons can convert to electron-positron pairs and then short conversion tracks appear in the tracker that because of the magnetic field spread in the ϕ direction. A dedicated algorithm based on multivariate techniques, uses the distance between ECAL clusters and their energy to cluster them together depending on how likely they are to belong to a converted photon cluster. Since the probability for the conversion of photons increases as the photon traverses more and more layers of tracking material, photons are expected to have conversion tracks that are not associated with a track pixel seed and this acts as a good discriminator between photons and electrons created promptly. More information about photon reconstruction and performance in the CMS detector can be found here [33].

3.1.4 Hadrons

Hadronized gluons or quarks, will generate showers of particles that leave track hits and calorimeter depositions. The PF algorithm will attempt to reconstruct all the individual particles from this process. Hadrons, if they are charged are expected to leave hits in the tracking material, energy in the ECAL and mainly the HCAL. Neutral hadrons, are not expected to leave any hits in the tracking material and only leave large energy depositions int the HCAL. So charged hadrons are reconstructed as associated tracks and energy clusters in the HCAL towers. If there is track pointing to a large HCAL energy cluster, one can test if the energy of the cluster is consistent with the track momentum. The difference between the HCAL energy and expected energy based on the track, can be attributed to neutral hadrons overlapping with charged hadrons. Thus neutral hadrons are either reconstructed as HCAL depositions that are not associated to tracks or as the residual energy in the HCAL that is not accounted by charged hadrons.

3.1.5 Muons

Muons are reconstructed in the detector [45], as tracks that originated from the interaction point and extend to the muon systems leaving hits there. Since the muon is a minimum ionizing particle, it is expected to pass through the detector and leave a small energy signature in the different detecting subsystems. Reconstructed tracks from the tracking detectors, are combined with the hits in the muon systems to measure the muon track properties, such as charge sign, momentum and directions.

3.2 Meta-Objects and variables

3.3 Jets

Jets are a phenomenological tool, particle physicist use to study the hadronization of color charged objects in the collider detectors. Jets are simply a collection of tracks, ECAL and

HCAL energy depositions. There are many definitions of what a jet is and that simply depends on what way you choose to look at the detector and group together an amount of energy from the different subdetector up to the HCAL. Jets in the past used to be clustered with cone algorithms that looked at high energy depositions in the HCAL and then define a cone and collect whatever lands in them. Today sequential clustering algorithms have emerged, that either cluster PF candidates together as in CMS or calorimeter depositions.

Jet clustering algorithms, use the momenta and angular distances between PF candidates and then merge them together. Clustering algorithms merge at each step two objects to for a new one. The two objects can either be PF candidates or simple products of the merging procedure. In this analysis, we utilize two main clustering algorithms. The first is the $anti-k_t$ algorithm and the second is the k_{T} algorithm [13, 14].

The clustering sequence for both algorithms used, decides if two objects should be merged together or not based on whether, the d_{ij} quantity defined below is smaller than the d_{ib} quantity 3.2.

$$d_{ij} = min\{{p_{\mathrm{T}}}_i^{2p}, {p_{\mathrm{T}}}_j^{2p}\}\frac{\Delta R_{ij}}{R} \tag{3.1}$$

$$d_{ib} = {p_{\mathrm{T}}}_i^{2p} \tag{3.2}$$

The indeces, i and j run over the two particle candidates to be merged. The ΔR_{ij} calculates the angular distance between them and the parameter R is known as the Radius parameter of the jet.

The difference between the $anti-k_t$ and k_{T} algorithms, is what value the parameter p takes and drives which objects get merged first, based on their p_{T}. For the $anti-k_t$ algorithm, p takes the value -1 while for k_{T} it takes the value 1.

Two jets collections are clustered from the event PF candidates. The first uses the $anti-k_t$ algorithm and a radius parameter of 0.8. This collection is referred to in this document

as AK8 jets. The second again uses the $anti-k_t$ algorithm with a radius parameter of 0.4 and the corresponding jets are labeled as AK4 jets. The utilization of the k_{T} algorithm in this analysis, is described in the Photon jets chapter.

The AK8 collection is filtered to have jets only with $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 2$. Jet quality selections are also applied to remove bad reconstructed muons or tracks or ECAL clusters that group together and fake a jet. The AK4 jets are required to have a minimum p_{T} of 50 GeV and an $|\eta| < 2.4$. The AK8 collection is a wide radius jet collection, with the capability of clumping objects together with angular distances up to 1.6 and thus is a good collection of jets to use for boosted final states with multiple objects merging together. AK4 jets on the other hand are used for the triggering procedure and not in the search it self directly. Both jet collections, are calibrated and corrected for detector reconstruction effects [36].

The $anti-k_t$ algorithm has a number of good features that make it one of the preferable algorithms for clustering jets in collider experiments. The most important of which is that the jets it creates have circular shapes, in the $\eta-\phi$ space. This is convenient when calibrating these objects unlike having jets with non-canonical shapes.

It is possible to calculate a likelihood for a reconstructed jet, of whether or not this jet contains any B mesons and thus the jet was initiated by a b quark. This process is known as b-tagging [44]. B mesons have long life times with respect to other hadrons expected in hadronization showers. The long life time translates to decay vertices that are away from the interaction point. The CMS tracking detector has enough resolution to discern these long-lived secondary vertices and thus identify b-meson candidates.

3.3.1 Event variables

The total hadronic activity (H_{T}) of an event is generally a sum of all the p_{T} of the jets in the event. Since there are multiple ways to define jets, and thus multiple collections of jets, there are also multiple definitions of the H_{T}.

In this search, we define H_{T} as the scalar sum of all the AK8 p_{T} in the event. The AK4 jets are used for triggering purposes.

$$H_{\mathrm{T}} = \sum_i |p_{\mathrm{T}i}| \tag{3.3}$$

where i runs over all the AK8 jets in the event.

The missing transverse energy is measured as the imbalance in the final objects momenta in the transverse plane. That is defined as the negative vector sum of transverse momenta of all the jets and other objects in the event.

3.3.2 Primary Vertex and Pile-up

Interactions of bunches of protons, create multiple scatters that the hard scatter event is embedded in. We define as the primary vertex, that is the interaction vertex of interest, the vertex that is associated with the largest transverse momentum. The specific definition is the square sum of the p_{T} measures of all the jets - in our case AK8 jets - associated with that specific interaction vertex. In another way of viewing this, we select the vertex that the most energy has flown in the transverse plane per bunch crossing and thus signifies the largest momentum transfer interaction.

The primary vertex, once identified, is then associated with all the reconstructed tracks in the event and final state objects, by studying the distance of the vertex to the reconstructed objects and their proximity. The PF candidate collection is thus cleaned from tracks and hadrons that seem to originate from pileup vertices. This techniques is known as Charged Hadron Subtraction(CHS).

Chapter 4

Data samples and Simulation

The search is performed on proton-proton collision data collected by the CMS experiment during the Run2 of the LHC. A number of SM processes are simulated for studies. The signal process is also simulated for the purpose of interpreting the results of the analysis.

4.1 Data

The data analyzed were collected by the CMS experiment at the LHC from proton-proton(pp) collisions at $\sqrt{s} = 13\,\text{TeV}$ during the 2016 data taking period, and correspond to an integrated luminosity of 35.9 fb^{-1}. A mask is applied on the event to reject runs that are problematic were known issues existed with detector. The datasets used in this search are listed in appendix A.

4.2 Triggering

Signal events, due to their large H_T, are expected to be selected by triggers that look for high H_T events. We used the JetHT data set, that is collected using a combination of H_T triggers, and apply a selection on the events, to pass a trigger requiring a minimum H_T of 900 GeV. The trigger H_T is calculated using AK4 jets with a minimum p_T of 50 GeV and $|\eta| < 2.5$. Triggers exhibit a turn-on curve, that is the cut is applied online and if you compared it to the offline selection on the H_T of the event, the two quantities are not identical and thus a turn-on is observed. Triggering systems also can exhibit inefficiencies either due to detector issues or readout and algorithmic issues. This can further decrease

the efficiency of a trigger by either increasing the turn-on range and even not allowing the trigger to reach a 100% efficiency, even well above the cut off value. In appendix B the studies performed to evaluate the efficiency loss in RunH are presented.

This was the case with RunH of the data collection. The H_{T} trigger was not reaching its 100%efficiency and to recover the lost events, a combination of jet based triggers is used. These include a jet trigger that looks for CaloJets reconstructed with a p_{T} greater than 500 GeV and a trigger looking for jets reconstructed from PF candidates, with a p_{T} of 450 GeV. A logical OR of the three triggers is used to filter events and recover trigger efficiency in RunH. In Appendix A the studies we performed for the trigger efficiency and recovery using the extra triggers is presented. The test of the trigger required using an orthogonal dataset, that is a data set collected with a totally independent way from the trigger that is to be studied. In our case, the Single Electron dataset, collected by triggers requiring the presence of single electrons in the events, is used to measure the efficiency of the jet H_{T} triggers.

For the purpose of extracting an efficiency correction for the signal, events were also collected with a combination of muon triggers, selecting events containing at least one muon with p_{T} greater than 50 GeV.

4.3 Simulation

Pair production of gluinos for a range of different $\tilde{\mathrm{g}}$ and $\tilde{\chi}_1^0$ masses, with the S and $\tilde{S}$ masses fixed to 90 and 100 GeV, respectively, are simulated using MadGraph _mc@nlo [6]. In fig. 4.1 the mass grid of the generated signal points is presented. The decay and hadronization is done with pythia [46] using the CUETP8M1 tune [35] for the underlying event and the NNPDF3.0 parton distribution functions (PDF) [42]. The detector simulation is done using the CMS fast simulation package (FastSIM) [3, 30]. To estimate systematic uncertainties related to the detector simulation, the full CMS detector simulation (FullSIM) based on Geant4 [4] is also used and compared to the FastSIM results. An uncertainty due to the hadronization model is evaluated by an alternative signal simulation with herwig

[8], and the TuneEE5c [31] underlying event tune. Signal events are normalized using the theoretical gluino pair production cross sections [12] at the next-to-leading order, and assuming a 100% branching fraction to the decay channel we study.

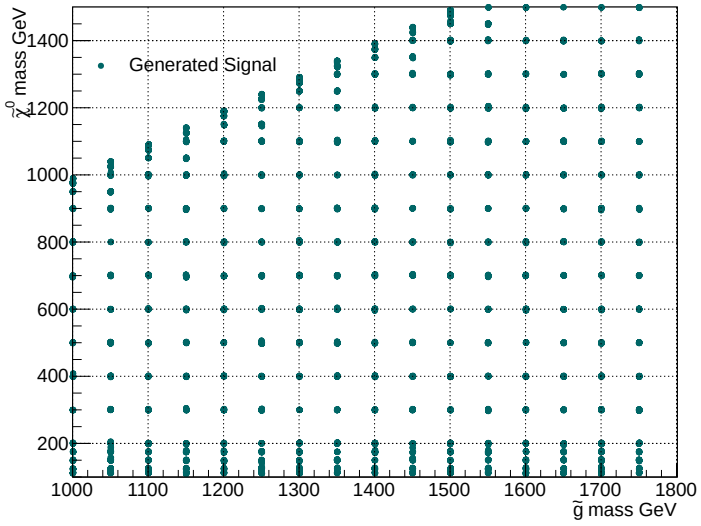

Figure 4.1: The signal mass phase space. The $\tilde{G}$ mass is taken to be zero, while the $S - \tilde{S}$ masses are fixed to 90 and 100 GeV each, for all mass points simulated.

We simulate SM processes to study the background behavior, to construct templates from which we estimate the efficiency corrections used for simulated signals, and to estimate the various uncertainties. The dominant source of background is quantum chromodynamics (QCD) multijet events. Simulation of QCD processes is done using MadGraph _mc@nlo with MLM matching [7] and hadronized with pythia8 with the CUETP8M1 tune. The production of hadronically and leptonically decaying W bosons in conjunction with jets (W+ jets) is simulated this way. Top quark pairs ($t\bar{t}$) are simulated with powheg v2 [41, 28, 5, 29] and hadronized by pythia8 using the CUETP8M2T4 [35]underlying event tune. As an alternative to pythia, the hadronization of $t\bar{t}$ pairs is also performed with herwig and the TuneEE5C underlying event tune. All samples are simulated with the NNPDF 3.0 parton distribution functions. The detector response is simulated with Geant4.

Chapter 5

Photon Jets

The target of this search is to discover possibly produced exotic states in the collider, by looking for events that have jets composed of a photon and two other cores. To identify these unique jets against QCD jets we follow a strategy, that aims to find the photon in the jet and the two cores coming from the gluons. Identifying photons in jets is no easy task. The photon is surrounded by the rest of the activity in the jet not allowing to utilize the photon isolation variables. Further more, the hadronization of gluons and quarks results in pions (π^0) that decay to photons. Pions coming from energetic mesons or other hadrons, will be Lorentz boosted and the daughter photons would merge to appear as single photons with higher momentum rendering them hard to distinguish from true final state photons originating from superparticles decays. The activity around the photons coming from pion decays is larger on average than final state photons outside jets and isolation selections help discriminate against them. QCD jets with a strong photon candidate in them are one of the background processes for this analysis that beyond the selections we apply to remove them, the background estimation method is based on quantifying their contribution in the signal search window correctly.

We develop a method to identify the subjets in the jet and then use them to construct a variable that can replace the isolation requirements on photons. Isolation criteria are usually a sum of the energy around the object- in this case the photon in a cone in the $\eta - \phi$ space. That means that the relevant information used for the isolation calculation, is a geometric distance between the photon and the other energy depositions around it. Our method is based on finding the subjets and then defining a new isolation using the subjets

p_{T}.

5.1 Photon Identification

Photon identification, refers to the process of examining reconstructed photon objects and determining if these objects are likely to represent a reconstructed true final state photon, or a a jet or other object, that manages to mimic the signature of photons in the detector. The goal of this process is to have a good efficiency of true photons combined with a strong fake rejection power. The decision is based on a number or features of the reconstructed object. These features can then be used to apply selections and thus define a label on what me refer to, when talking about good photons reconstructed in the detector. There are two paths one can take in this approach. The first - and the one we follow - is to derive a set of selections (cuts) on these variables that are related to the reconstructed object, and define thus a label of what we mean by a good photon. This method is referred to as a Cut Based method. The second approach is to take the observable features and use them to train a Multivariate discriminator that returns a single score and classifies the objects into true photons and possible other objects that fake it. Since the beginning of my PhD I have derived for the CMS experiment multiple Cut based Identities for photons that have been used by other analysts to perform searches. This knowledge was vital for the completion of this project.

The variables that one can use to select to photons, are related to the cluster of energy in the ECAL, the photon has deposited. Further more the presence or not of a good track associated with this cluster of energy, can be used as a discriminator between photons and electrons, since electrons leave a good track in the tracker before they shower in the calorimeter. The standard photon Cut Based ID of CMS uses 5 variables to reject jets and other objects that mimic the photon, plus one for discriminating between photons and electrons.

The first variable is a ratio of the energy in the HCAL towers directly behind the ECAL

energy cluster associated to the photon, to the energy present in the ECAL and belonging to the photon cluster. This ratio that we refer as H/E (H for hadronic Energy and E for the Electromagnetic energy), manages to rejects early showering hadrons in the ECAL that can leave a considerable ECAL energy. Especially neutral hadrons will result in no track associated to the ECAL cluster and their energy shared between the two calorimeters, making the HCAL energy information critical for rejecting them. High energetic photons will in principle leak energy in the HCAL towers behind the ECAL crystals, and thus it is necessary to understand which is the energy range of photons of interest as to select an H/E cut that would optimize against neutral and other hadrons as well as not killing photons that leaked in the HCAL. The H/E distribution of signal photons is presented in fig. 5.1 , compared to object originating from QCD jets and are reconstructed as photons. In this search and given that the photons are embeded in a jet, photons are required to have less than 5% of their total energy deposited in the HCAL towers behind them.

The second variable that relates to the energy deposition in the ECAL, is the $\sigma_{i\eta i\eta}$ variable. The variable is define in eq. 5.1.

$$\sigma_{i\eta i\eta} = \frac{\sum\limits_{i} w_i (i\eta - i\bar{\eta})^2 {d_0}^2}{\sum\limits_{i} w_i} \tag{5.1}$$

$$i\bar{\eta} = \frac{\sum\limits_{i} E_i i\eta}{\sum\limits_{i} E_i} \tag{5.2}$$

$$w_i = max(0.0, w_0 + log\frac{E_i}{E_{5x5}}) \tag{5.3}$$

where $i\eta$ is the index of the i^{ith} crystal in the cluster, E_i the energy in that crystal, E_{5x5} the total energy in a matrix of 5 by 5 crystals centered on the highest energy crystal in the supercluster, d_0 is the size of individual crystals in the η direction and w_0 a constant that is used to kill off crystals that have lees than 0.9% of the total 5x5 crystal matrix. This

seemingly complicated expression is nothing but a weighted standard deviation calculation of the energy spread in the crystal space along the η direction for photon supercluster.

The distribution of this variable, is presented in fig. 5.1. Signal photons are compared to photons from other sources. The variable is also split for Barrel photons and End Cap photons separately since the crystal configurations of the ECAL is different for the two cases, as explained before.

There are three more variables that are usually utilized in a Cut Based ID for photons. These are three flavors of isolations variables. Each measures the activity around the photon objects, looking at different PF objets, either other PF photons, neutral hadrons or charged hadrons. The isolation variable is defined as the sum of the p_{T} of the PF objects of a specific family in a cone around the photon with a radius in the $\eta - \phi$ space of 0.3. But this is not a simple matter.

The problem lies in the fact, that photons can convert in the tracking material in-front of the calorimeters, resulting in a shower of particles (electrons and photons) being deposited in the crystals. The PF candidates around the photon are thus needed to be cleaned from the photon footprint - as we refer to these shower components from conversions - to not overestimate the isolation around the photon and thus reject it as fake. These requirements can not be utilized in our analysis, since the jet activity spoils them and makes it hard to use to discriminate between true final state photons and photons coming from pion decays or fakes. The photon variables are presented in fig. 5.1 and compared for photon coming from the signal process and photons in data and from simulated SM processes.

The other important variable for photons is the pixel seed. This variable flags if there is a pixel seed that a track from it points to the ECAL energy cluster associated with the photon. The pixel seed is set of three hits in the pixel layer that act as a seed for track reconstruction. ERlectron produced promptly in the detector, will have such a set of hits in the pixel detector. Photons on the other hand that propagate with out converting through the detector, will not have such a set of hits associated with them. If the photons convert

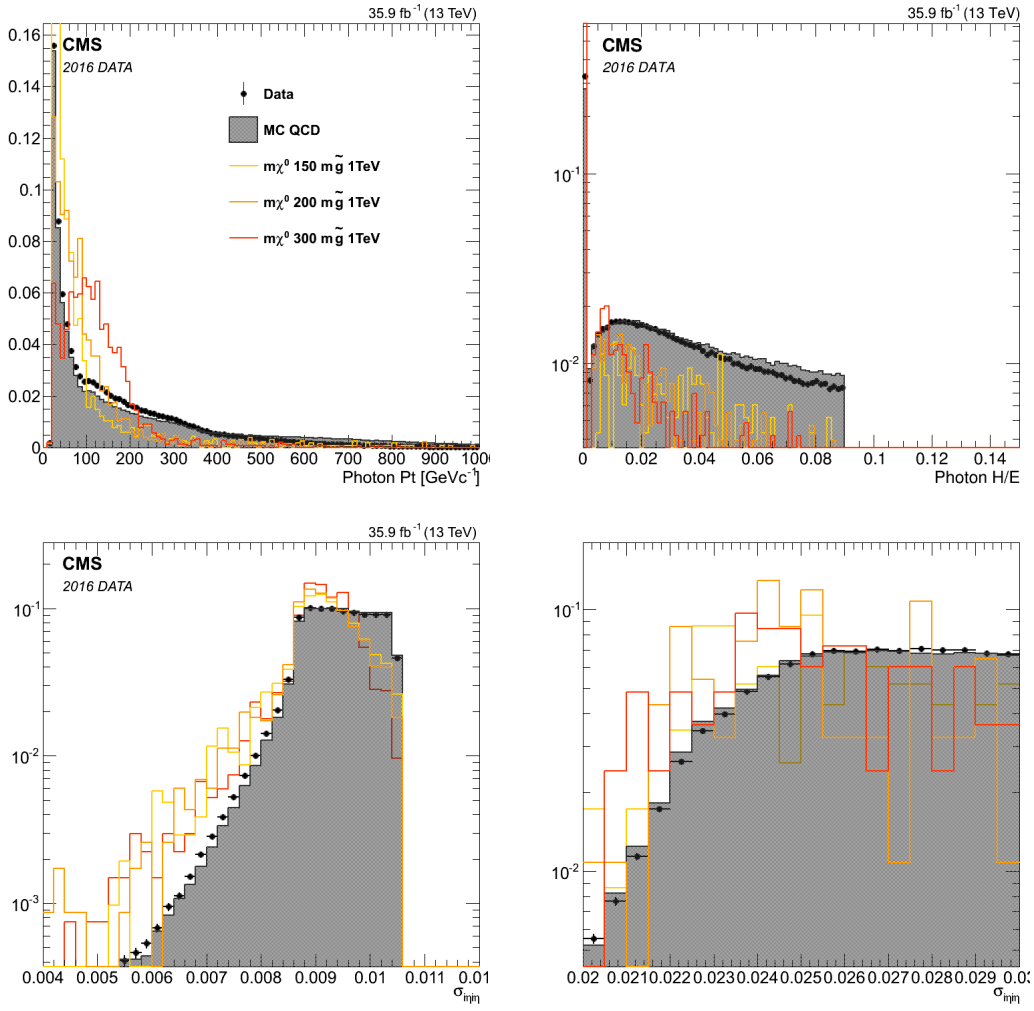

Figure 5.1: Normalized distributions for the photon variables. Signal, data and QCD simulated photon distributions are presented. The p_{T}, H/E and $\sigma_{i\eta i\eta}$ variables. The $\sigma_{i\eta i\eta}$ is split into Barrel photons (lower left) and EndCap photons (lower right)

though, they will have conversion tracks associated with them. The probability for the photon to convert, increases as we move radially out from the interaction point since the tracking material increases. So lees converted photons are expected to have converted in the early pixel layers and thus generating this set of hits in the pixel detector. This variable is used as a good discriminator between photons and electrons, and we use it to measure the scale factor for our analysis.

5.2 Finding the subjets

We develop an algorithm to identify the three cores in the jet, one of which is the photon. To do so, we look at the merging history of the jets under examination and identify as subjets the last three objects that merged to form the final jet. Jet clustering algorithms, take as input a collection of four-vectors and sequentially merge them to form a final jet. At each step of the clustering sequence two objects are merged to form a new one. Different clustering algorithms use different metrics to decide which objects are to be merged and not and with which order. The order does not affect the final jet properties but it affects the the subjet definitions, if these are identified as intermediate steps in the clustering sequence. There are three algorithms that we consider. The $anti-k_t$, k_{T} and Cambridge-Aachen clustering algorithms. The metrics that the algorithms use are a combination of the angular distances between the objects to be clustered and also their momenta. The Cambridge-Aachen (CA) algorithm uses only the angular distance as a criterion to what to merge first, while the kt algorithm, merges first that are soft and close in the $\eta-\phi$ space, while the $anti-k_t$ algorithm merges first the more energetic objects together.

5.2.1 Photon Jets Algorithm

First, we check if the AK8 jets that we consider, have one or more reconstructed photons in them , that is with an angular distance less than 0.8 between the photon and the jet. These reconstructed photons are demanded to have minimum p_{T} of 20 GeV. Following

that, the collection of the PF candidates that composes the jet, is cleaned from the photon footprint. The reconstructed photon replaces then the removed objects, and the clustering sequence is run on the pf candidate collection. Not removing the photon footprint from the PF candidate collection, can result in part of the photon leaking in other subjet than the one it should be and thus appearing less energetic with respect to the subjet it is a part of.

After reclustering the jet, the merging history is then examined. The two objects that merged to form the final jet, are taken. The lighter of the two in mass, is taken as the first subjet of the jet. The other one, the heavier of the two, is examined further to see what objects merged to form it. The two objects that merged to form it are then considered the 2^{nd} and 3^{rd} subjets of the jet. We demand that 3 such subjets are found in the jet each with a minimum p_{T} of 10 GeV. In fig. 5.2, the three subjet p_{T} distributions, returned by our algorithm are presented.

5.3 The photon subjet fraction (f_γ)

Once we have the subjets, we check in which of them the photon ended up grouped in. The ratio between the photon p_{T} and the subjet p_{T} is then calculated. This is defined as the photon subjet fraction (f_γ) 5.4. In the case multiple photons are found in the jet, all the f_γ are calculated for each and the one with the largest value is selected to be the f_γ value for the AK8 jet. In fig. 5.3 the distribution of f_γ for signal jets is presented as for data and QCD jets. Signal jets, tend to have values close to one, while QCD jets are spread uniformly with a drop at the high values close to one.

$$f_\gamma = \frac{p_{\mathrm{T}\gamma}}{p_{\mathrm{T}\,subjet}} \tag{5.4}$$

This variable replaces the photon isolation variables and can be used to reject candidate jets as Photon Jets. One way of viewing the clustering sequence into subjets and then calculating the photon isolation in the form of the f_γ, is that this metric that is used for

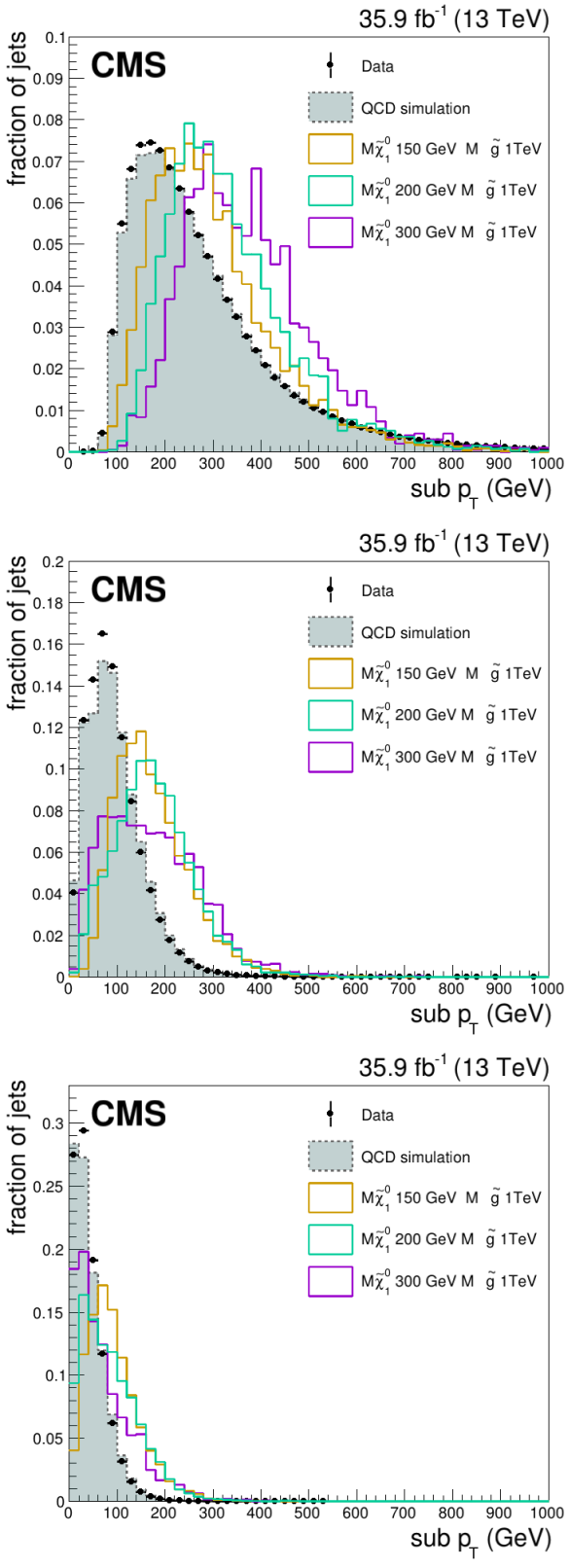

Figure 5.2: The normalized subjet transverse momenta, reconstructed by our algorithm. From top to bottom the leading, subleading and third leading subjet p_{T}. The signal distributions here correspond to matched signal jets, that is simulated signal jets that contain the photon and the two gluons of the singlet decay.

the calculation of the the activity around the photon does not only incorporate the angular information between the energy deposits and the photon as in the simple isolation variables case. Instead their energy is also incorporated in the decision of including or not an object in the total activity. Thus in a more complicated fashion, based on the clustering algorithm that is selected, near-by object that would have been counted in the isolation cone can be separated in different subjets.

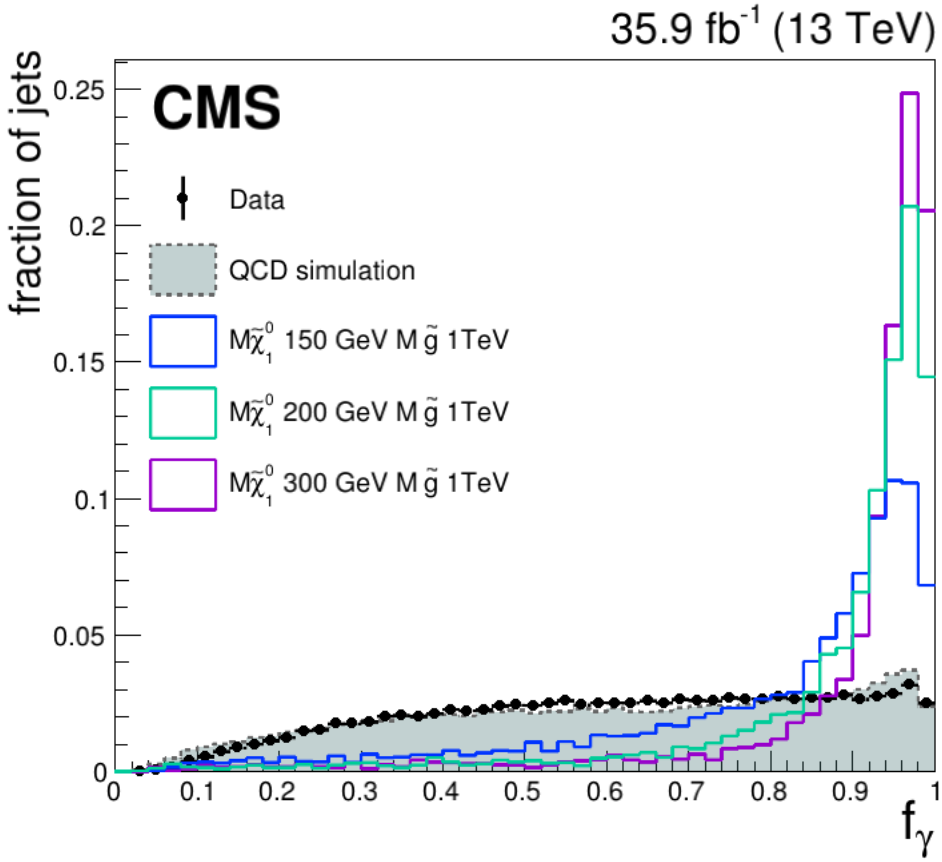

Figure 5.3: Normalized f_γ variable. Signal distributions here correspond to simulated jets that contain the neutralino decay products. Signal jets, tend to have a high f_γ value while QCD jets are expected to be spread all over. By narrowing on the region the signal accumulates, it is possible to reject the majority of the QCD background.

This algorithm is run over all AK8 jets in the event that fulfill the kinematic criteria. For each of them that the algorithm returns three subjets with $p_{\mathrm{T}} > 10$ GeV, the subjet fraction is calculated. The selected reconstructed photon is also marked for each jet to be

tested further.

5.4 N-subjettiness

There is another useful tool in identifying substructure in jets, that does not require identifying individual subjets. These are the N-subjetiness variables [51]. These variable take the PF candidates of each event, and calculate a number between 0 and 1 that signifies how probable it the jet to have N subjets in it. The more likely there are N-subjets in the jet, the lower the value of this variable. First N axis are defined in the jet. Then the pt of the PF candidates is multiplied with the min angular distance between it and the closest axis. A sum is taken of all these quantities for each PF candidate and normalized.The variable is defined in eq. 5.5.

$$\tau_N = \frac{1}{d_o} \sum_i p_{\mathrm{T}i} min\{\Delta R_{1,i}, \Delta R_{2,i}, ...\Delta R_{N,i}\} \tag{5.5}$$

where:

$$d_0 = \sum i p_{\mathrm{T}i} R_0 \tag{5.6}$$

The index i in the sum runs over all four vectors that are clustered to make the jet. The R_0 variable is the radius parameter used to cluster the jet and d_0 acts as a normalization factor restricting the variable between zero and one. The expression in the brackets, denotes that one has to select the minimum angular distance between the four vector i and the N axes that have been defined in the jet to calculate N-subjetiness. These variables can be used to test different possible number of axes-subjets in the jet and they reflect how the energy is distributed in the jet, if it collimates around one or two or three axes. Taking ratios of these variables for example τ_2/τ_1, can prove a test between two hypotheses, that is between a 2 core jet vs a single core jet, since if the jet is more 2 prong like, will have a lower τ_2 value and a higher τ_1 value. As a result the ratio of the two would be small and it provides a nice way to package together the two variables and apply a single cut that

ensures discrimination between 1-prong and 2-prong jets.

We are trying to identify jets that have a 3 prong structure. Thus we utilize the τ_3/τ_1 ratio to separate signal jets from SM jets. The distribution of the this variable is presented in fig.5.4. Signal accumulates in small values closer to zero, while QCD jets are shifted in higher values with a tail extending in the low values too.

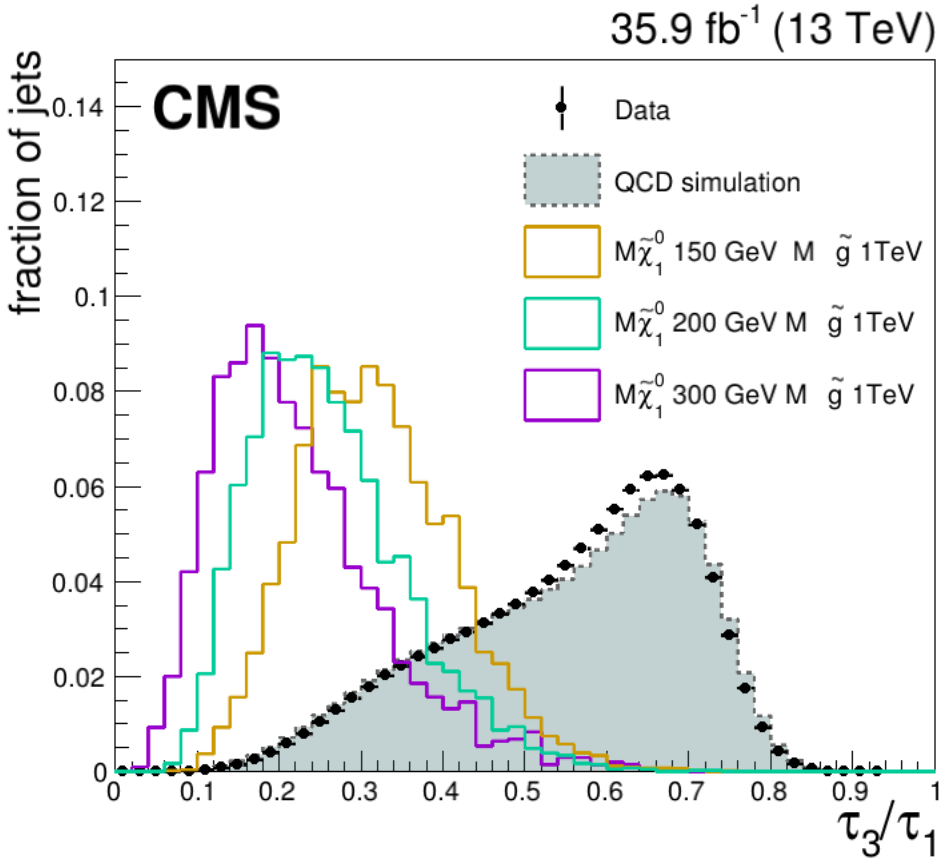

Figure 5.4: Normalized distribution of N subjetinness τ_3 over N subjetinnes τ_1. This ratio is a good discriminating variable between jet that a possess 3 prong structure vs a single prong jets. The plot shows all jets in signal events matched to neutralino decays next to jets from data and simulated QCD events.

Having in our toolbox this information for each of the jets, we proceed with designing labels and selections to create a search window and discover possible new phenomena hiding in the large SM background processes at the LHC.

Chapter 6

Analysis Design

Following the construction of the f_γ variable in the previous chapter and the presentation of all the relevant variable related to jets and photons as well as events, in this chapter we outline how everything comes together to the design of the analysis, with the different categories labels and the definitions of the background dominated and signal search regions.

6.1 Event Selections

As it was outlined before, we target multijet events with large hadronic activity. Because of this, we apply an $H_{\mathrm{T}} > 1\,\mathrm{TeV}$ cut on the events under consideration, that suppresses QCD contributions and other SM processes . The H_{T} for signal events is expected to be on the order of the mass and more of the produced resonance, that is the mass of the gluino, since all the final states are observable deposit energy in the detector.

The major SM background contribution high jet multiplicity events with large hadronic activity, are QCD processes. In fig. 6.1 the distribution of the number of AK8 jets per event is presented for signal events and SM qcd events. QCD tends to be characterized by di-jet events with a large tail towards greater multiplicities of jets. Signal events on the other hand, have greater number of jets per event .

We create two categories of events to perform our search. The first, is the set of events that have exactly three AK8 jets and the second is the set of event with four or more AK8 jets. This way, we create one region that is more sensitive than if we would merge the two sets into one and perform the search since the QCD contribution to the expected number

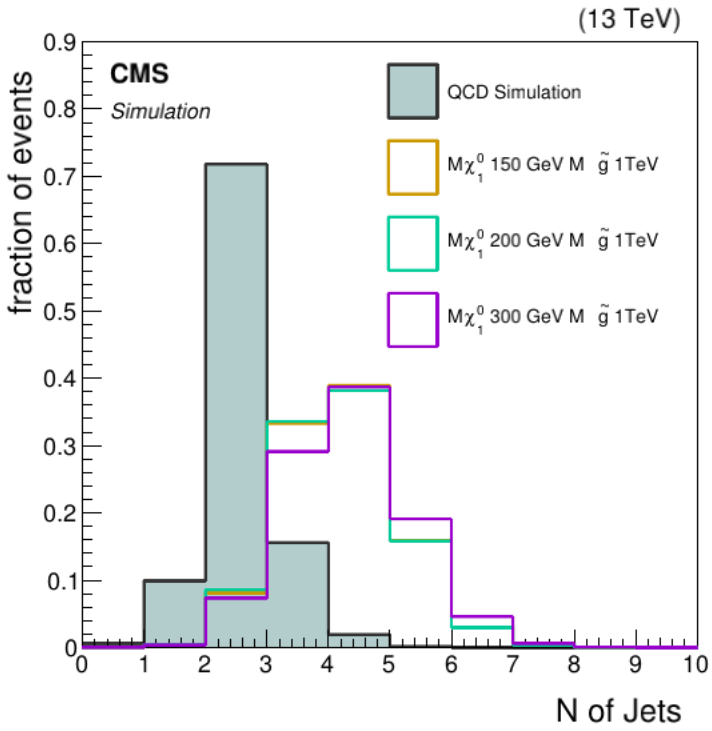

Figure 6.1: Distribution of total number of AK8 jets with pt $>$ 200 GeV per event. QCD events are largely di-jets events while the bulk of the signal events tend to have more than 2 jets per event.

of signal events for the 4+ AK8 jets category is smaller than in the 3 jets category. We have thus created two sets of events for the search to be performed on. Both with an H_{T} $> 1\,\mathrm{TeV}$ requirement and then the two set split by the number of AK8 jets they contain.

6.2 Definitions and Selections

The next step, is to utilize the photon and jet variables we constructed and presented to create labels for unique jets that are signal like. We would like to look at the AK8 jets in each event and count how many unusual jets exist in each event that appear to have substructure and a strong photon component in them. These unusual jets, are defined to be *photon jets*. We create two categories of photon jets, one with looser selections and one with tighter selections.

Loose jets, are defined as the AK8 jets that have a three subjets in them with $p_{\mathrm{T}} > 10$ GeV, a good photon candidate in them and a $\tau_3/\tau 1 < 0.4$. A good photon candidate is defined as a photon with a $p_{\mathrm{T}} > 20$ GeV, an $H/E < 0.05$ and a selection on the $\sigma_{i\eta i\eta}$ < 0.0105 for Barrel photons and $\sigma_{i\eta i\eta} < 0.0305$ for EndCap photons. The photon is also required to not have a pixel seed associated with it. These selections, ensure that a Loose photon jet, will be composed of three cores one of which is a good photon candidate. As Tight photon jets, we define Loose photon jets, that also are required to have $f_\gamma > 0.9$. Tight photon jets are thus a subset of the Loose photon jets. This way we ensure that the photon found in the jet, is actually one of the three subjets and dominates it. The definitions of Loose and Tight photon jets, are presented in the table 6.1.

6.3 Search Regions

With the Loose and Tight photon jets definitions at hand, we can label individual events based on the number of Loose and Tight jets they posses. We develop a naming scheme that is used through out the thesis and the search. An event with X jets passing the Loose photon jet selections, of which Y also pass the Tight photon jet selections, is labeled as

variable	Loose Photon AK8CHS Jets	Tight Photon AK8CHS Jets
Jet Pt	> 200 GeV	
Jet $\|\eta\|$	< 2.4	
τ_3/τ_1	< 0.4	
photon	has good photon	
subjets	has 3 subjets with min pt > 10 GeV	
f_γ	no selection	> 0.9

Table 6.1: Loose and Tight photon jets definitions. The integrated efficiency for signal jets with χ_1^0 mass of 200 GeV and gluino mass of 1TeV is about 66% for the Loose selections and 45% for Tight, while the measured Data mistag rate is 1% Loose and 20% Tight.

Figure 6.2: Cut flow table for signal matched jets and different mass points.

Event Signal Region Definition
$H_T > 1$TeV
3 XOR 4 plus AK8 jets Jets min ΔR of 1.5
2 Loose Photon Jets as defined in 6.1
Further split SR in bins of N Tight Photon jets

Table 6.2: This table summarizes the Signal Regions definition. Events are required to satisfy the total hadronic reqirement, have 3 or more AK8 jets and have 2 Loose jets as to be consider belonging to the Signal search region. These selections greatly reduce the SM processes that contribute in the SR.

an LT X-Y event (LT stands for Loose Tight). That is an event labeled as LT 21 means that the event has 2 Loose photon jets of which 1 is also a Tight photon jet, while an event labeled as LT10, means that this event has 1 Loose photon jet only that is not a Tight photon jet. The number of Tight photon jets is always less equal to the number of Loose photon jets, since they are a subset of the Loose jets. Thus the possible event categories for LT2X are, LT20, LT21 and LT22, which refers to events with zero, one and two tight photon jets respectively.

Signal events beyond the large hadronic activity and multiple AK8 jets, are expected to have two of the unusual jets composed of the photon and the two gluons from the neutralino decays. This means that we can expect that the signal events will have two jets that would pass the Loose photon jet selections. We define the search window , or signal region (SR) as it is known in the particle physics community, the set of event that have at least two Loose photon jets, that is LT 2X. The selection defining the signal region, are all listed in table 6.2.

Events coming from SM processes, are expected to dominate the set of events that have less than two Loose photon jets. QCD jets will emit final state gluons suppressed by a_s. Then also the probability for objects in the QCD jet to be labeled as a photon is of a small probability. So having two jets that both these conditions are fulfilled becomes a relatively rare occurrence with respect to the totality of QCD processes in the collider. There is no reason why two jets in QCD events should on average be labeled as Loose photon jets with

a high probability, unlike for signal event.

We thus construct based on these event labels, a space where background and signal events live. In fig. 6.3 this space is presented. The ordinate in this figure counts how many Loose photon jets exist in the event and the abscissa the number of Tight photon jets. Events populate the region above the diagonal since as mentioned above Tight jets are a subset of Loose photon jets. The reason we split events also in how many Tight photon jets they contain, is that the more Tight jets the event contains, the more sensitive it is to new physics, with signal events dominating the background events. The H_{T} distributions of data are compared for each one of the SR categories, to the expected H_{T} shapes, originating form SM processes, to discover potential new states.

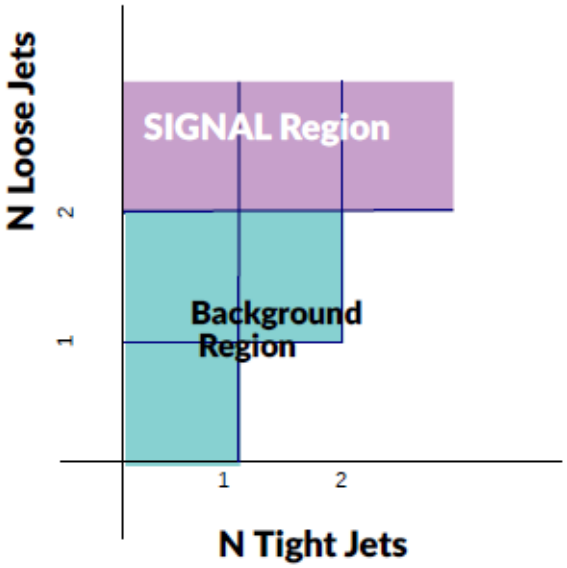

Figure 6.3: The SR and BR based on the number of Loose and Tight photon jets

Chapter 7

Background Estimation

Background estimation refers to the procedure analysts use to estimate the background processes contribution in the search windows. This can be either a shape or an event count or both. It is necessary to have a method that can provide reliable background predictions with known uncertainties. For this reason, the methods are designed and tested for closure,that is the capacity of the method to predict the background shapes and event yields correctly. In this chapter we describe the method we developed and used to estimate the background H_{T} shapes and event yields in this analysis.

7.1 Ensemble method

The main background component in the signal region is QCD multijet events. Using simulated events, we also estimated the $\mathrm{t\bar{t}}$ and W+jets contribution in the signal region categories and they were found to be of the order of a few percent compared to the QCD contribution and thus neglected further in this analysis. We thus focus on predicting correctly the multijet QCD background. The main idea of the method we use is to measuring in data the probability for a generic QCD jet to be labeled either Loose or Tight photon jet. Then using this mistag rate to take the events in the background dominated region and for each calculate a probability to populate any of the signal region categories. The background expected H_{T} distributions for the signal region , can then be constructed by plotting the H_{T} of all the events in the background region and weighting them with the probability that we calculated from the mistag rates. The H_{T} distributions for the expected SM process in each of the 3 signal region bins for either the 3 or 4 or more AK8 jets, can be obtained by

changing the weights to the probabilities that correspond to each respective signal bin.We assume that the mistag rates depend uniquely on the jet kinematic properties, that is the jet p_{T} and η. To calculate the event probability for an event then to be propagated in the signal region, is just the product of the individual jet mistag rates, based on each of the event's jet's p_{T} and η arranged accordingly to which signal region bin the event is calculated to propagate into.

The event probabilities can be calculated analytically but the calculation becomes convoluted if one considers all the possible combinations for the signal region bins. Instead one can also utilize pseudo experiments, avoiding thus writing the expressions for the probabilities. This method begins by taking each event in the background region and then for each jet in the event, a random number is compared to the mistag rate for that specific jet (based on the jet's kinematics). If the random number is less than the mistag rate, the jet is considered to be labeled as Loose or Tight Photon jet considering which mistag rate is compared to. Each pseudo experiment provides a classification of the event under consideration and an ensemble of pseudo experiments, converges at the true probability value. For each event 10^4 pseudo experiments are performed. An interesting feature of this method, is that the mistag rates, allow to calculate a probability for each event, not only to populate the signal region but also the background dominated region. This feature as will be explained later, is used to test and validate the method.

7.1.1 Mistag Rates

The method is based on data measured mistag rates and the data events in the background region. We want to measure the mistag rates as a function of the jet p_{T} and η. The p_{T}-η distribution of jets passing the Loose photon jet category belonging to event in the LT 1-0 and LT 1-1 categories, is divided to the distribution of all the jets in the background region. The result is the data measured loose photon jet mistag rate. For the Tight photon jet mistag rate, the p_{T}-η distribution of Tight photon jets from the LT 1-1 events category, is

divided with the distribution of loose photon jets from the LT 1-0 and LT 1-1 categories. The Loose and Tight mistag rates are plotted against the jet p_{T} in fig. 7.1.

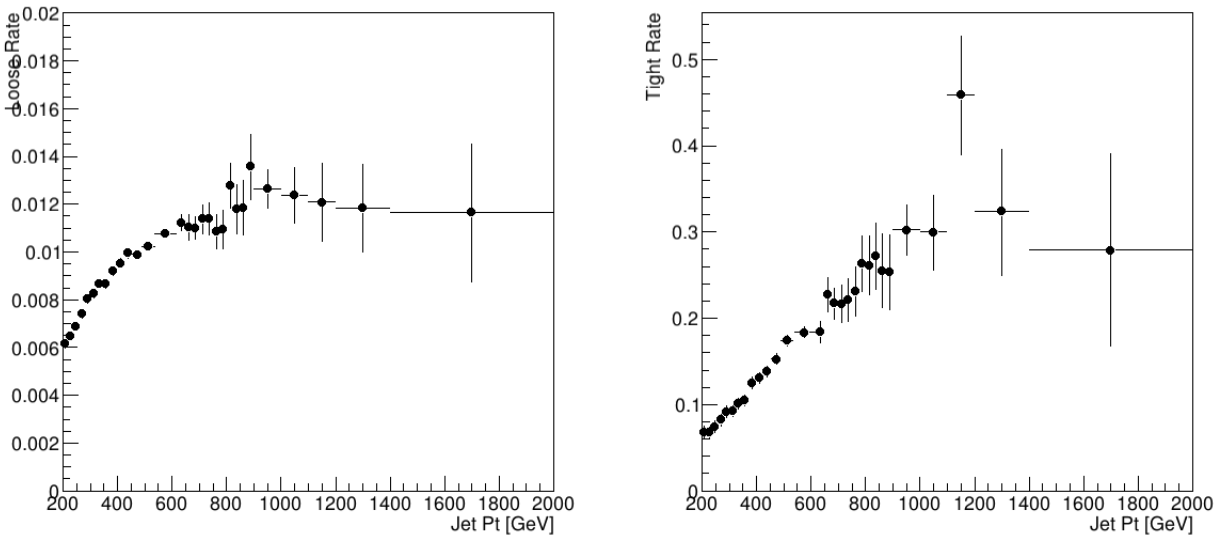

Figure 7.1: Loose (left) and Tight (right) mistag rates vs jet p_{T} as measured in the background dominated in data. The binning presented here is the one used to split jets and then fit the η distribution in each of these bins.

After measuring the mistag rates, we perform a fit on the rates to parametrize them with respect to the kinematic variables. A one dimensional fit (in η) is performed for the rates in p_{T} bins. The fit function used is a second order polynomial. The fact that the method predicts how H_{T} distribution should look like in the background dominated region, is used to optimize the mistag rates parametrization. In fig. 7.2 the fitted rates for a number of jet p_{T} bins are presented. The main uncertainty source on the rates is statistics, which then is propagated to the fit results are uncertainties on the fit parameters. The fit functions are used as input to the pseudo experiments method to calculated for each jet the mistag rate. The uncertainties on the fit parameters, are used to shift the mistag rates up and down propagating the uncertainty in the final result of the background estimation, the expected background H_{T} shapes in the signal region.

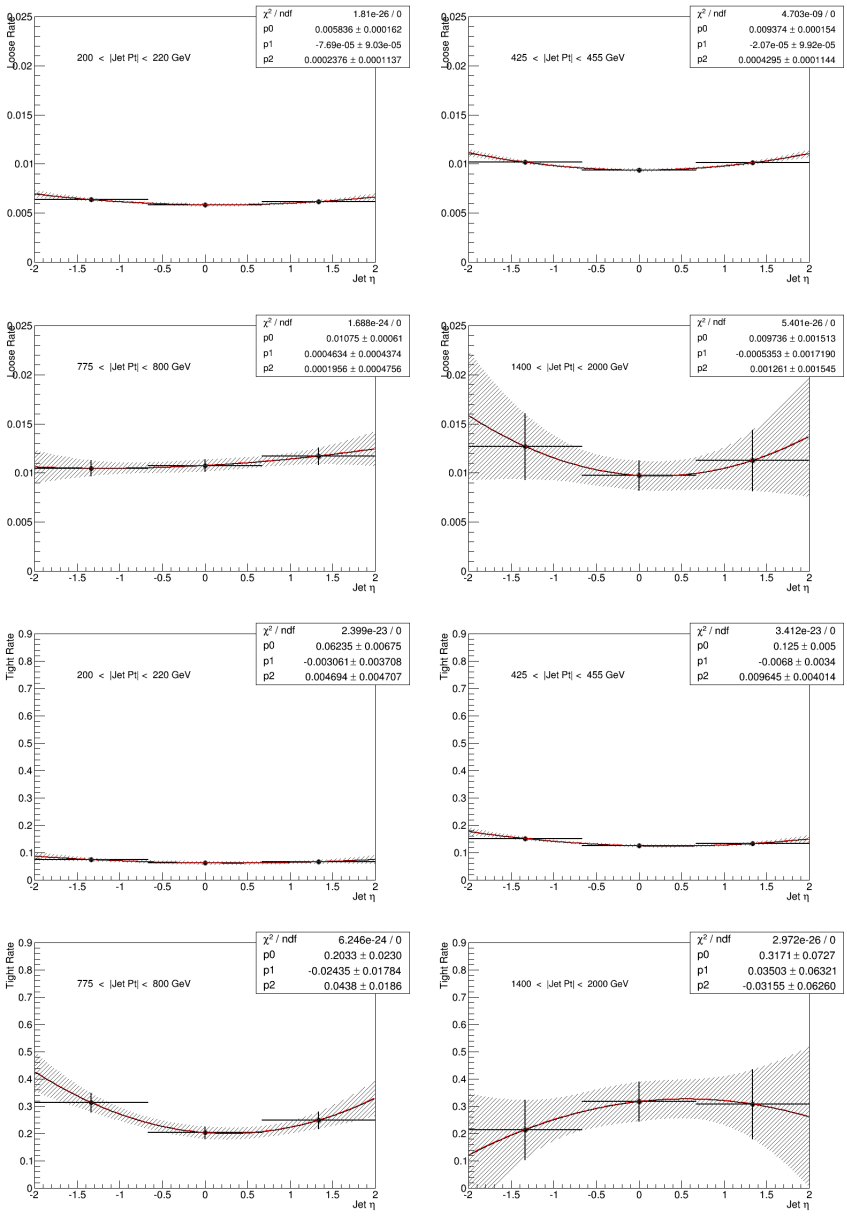

Figure 7.2: Loose and Tight Rates vs η for different pt bins as measured from the BR of the full dataset available. The fits for each bin are presented. A 2^{nd} order polynomial is used to fit the different bins. The gray band indicates the error on the mistag rates, that is used and propagated to the background estimation result. The error is driven from the statistics in each p_{T} bin.

7.1.2 Validation - Closure tests

To test the method's capacity to correctly describe the background process contribution in the signal region, the method is implemented on QCD simulated multijet events. The results of the background prediction, are compared to the results of applying the signal selections on the QCD simulation. The H_{T} shapes and event yields in the signal region, as well the background dominated regions are compared for consistency.

In the first casts of the method, the ensemble method, was underestimating the background contributions in the signal region. We examined a number of effects that could explain this result. The mistag rates were found to not vary considerably with the pile-up of the event. Different parametrization of the mistag rates were also used to test if the expectation and selection results would match. These attempts provided no remedy to the problem.

Based on that we questioned our basic assumption for the mistag rates, that is that the rates depend uniquely on the jet kinematics. Activity in the event, beyond pile-up could affect the mistag rates. We found out that the culprit was related to the jet clustering algorithms behavior in certain conditions. If two jets were too close to each other in the $\eta - \phi$ space, the algorithm favors one of the two jets, aggressively clustering components of the low energy jet as components of the high energy one. This results, in jets, appearing to be substructure rich and thus passing our signal selections. No information about jet proximity is incorporated in the ensemble method and in this cases the assumption we made fails.

To further test this, we applied a minimum ΔR between jets for each event and the ensemble method correctly predicted the signal selection results. In fig. 7.3 the disagreement between the cuts result and the ensemble method, is plotted against the minimum ΔR of jets. The disagreement appears for values below 1.5 and thus an extra selection on all events of a $\Delta R \geqslant 1.5$ is applied to ensure that the background estimation method functions properly. A test for the event yields using QCD simulated events, is presented in fig. 7.1.

Bin	Ensemble Prediction	Expected
	3 jets category	
LT20	2168.6 ± 346.3	2146.5
LT21	443.4 ± 97.0	468.3
LT22	23.4 ± 6.2	25.0

Table 7.1: This table presents total yields obtained with the ensemble method for different SR bins compared to the expected number, using QCD simulated samples. The numbers serve only to demonstrate the efficacy of the method to correctly predict the background contributions in the SR. We observe that the total yields predicted by the background estimation method and the expected numbers agree. This property is what is described as *closure*.

The predicted H_{T} distributions using QCD simluated events are compared to the results from applying the analysis on the simulated events in figures 7.4 and 7.5. The method is tested on a 10% subset of the total dataset. The results are presented in figures 7.6 and 7.7.

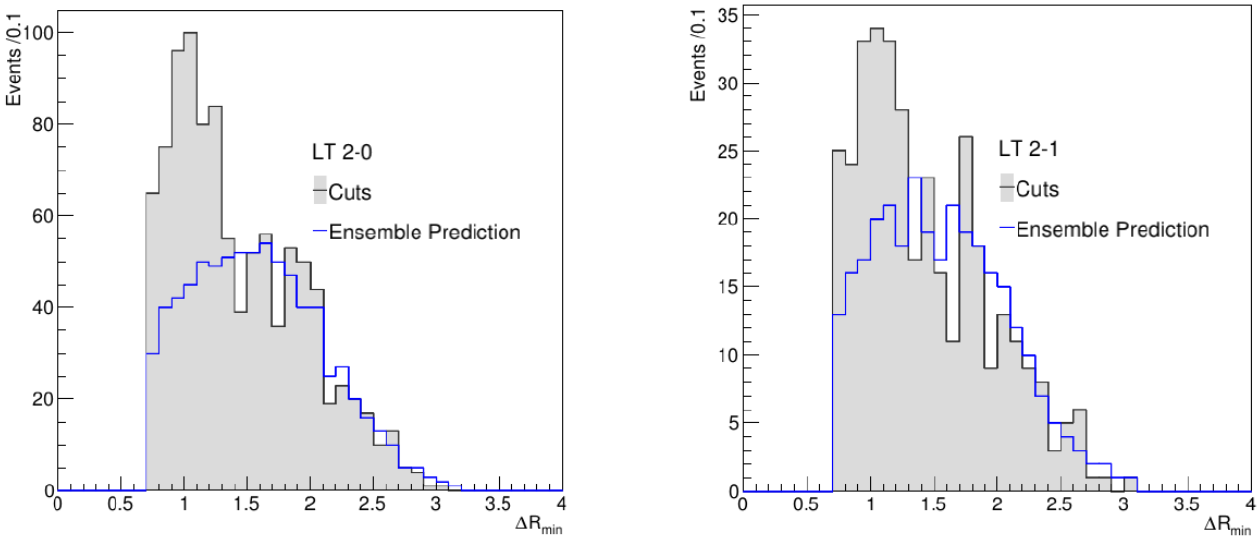

Figure 7.3: The estimated event counts vs the min ΔR between AK8 jets in the event. The gray area represents the result if one applies the SR selections while the blue line shows what the background estimation method predicts. It is clear that ensemble method (the background estimation method that is) does not agree with the selection results for event that the min ΔR is less than 1.5.

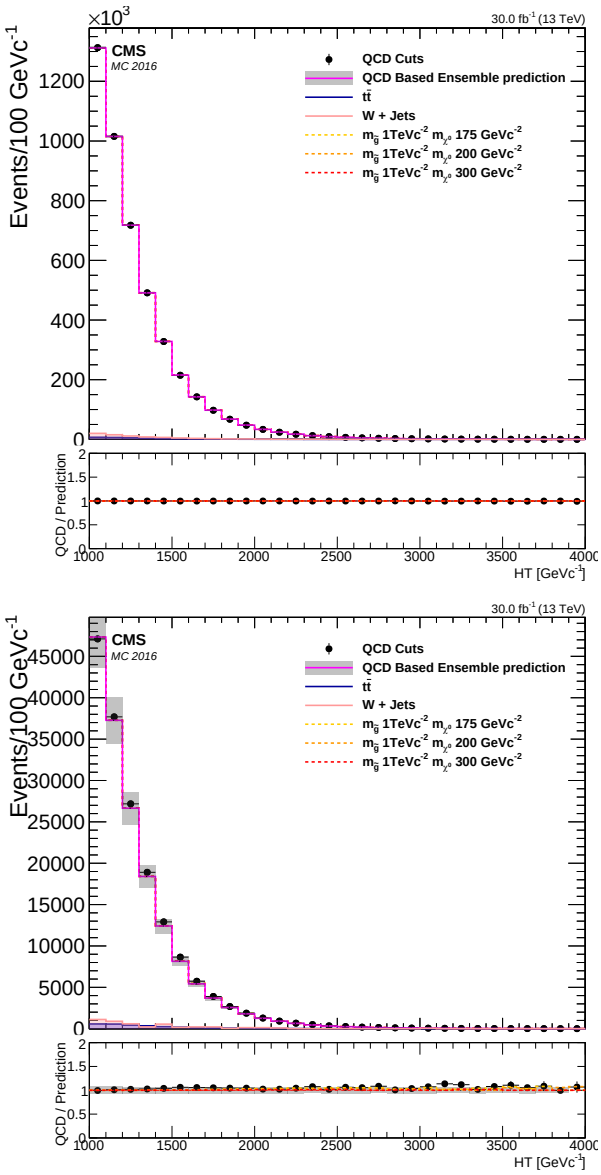

Figure 7.4: H_T distribution of 3 jet events having nor Loose or Tight jets LT(00) (top) and H_T distribution of 3 jet events having a single Loose jet which is not a Tight one LT(10) (bottom)

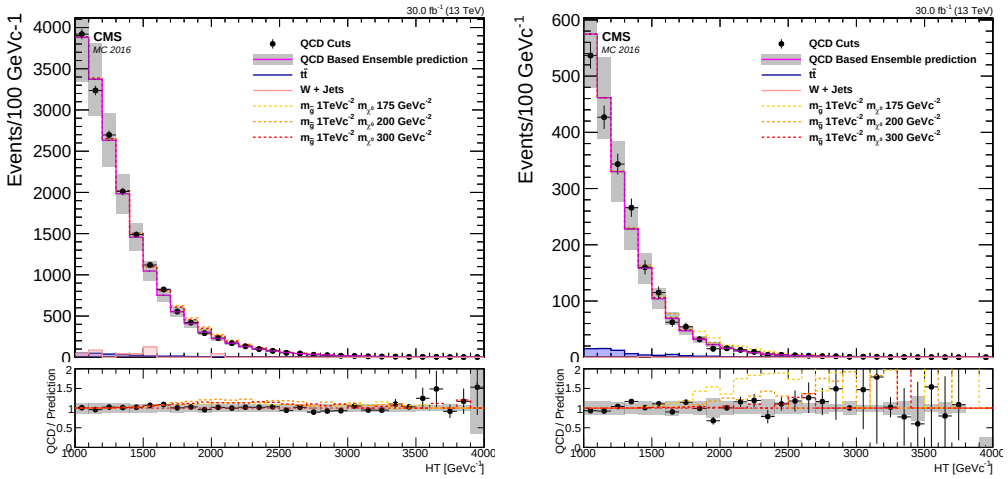

Figure 7.5: H_{T} distribution of 3 jet events in the LT 11 bin (top) and the LT 20 bin (bottom).

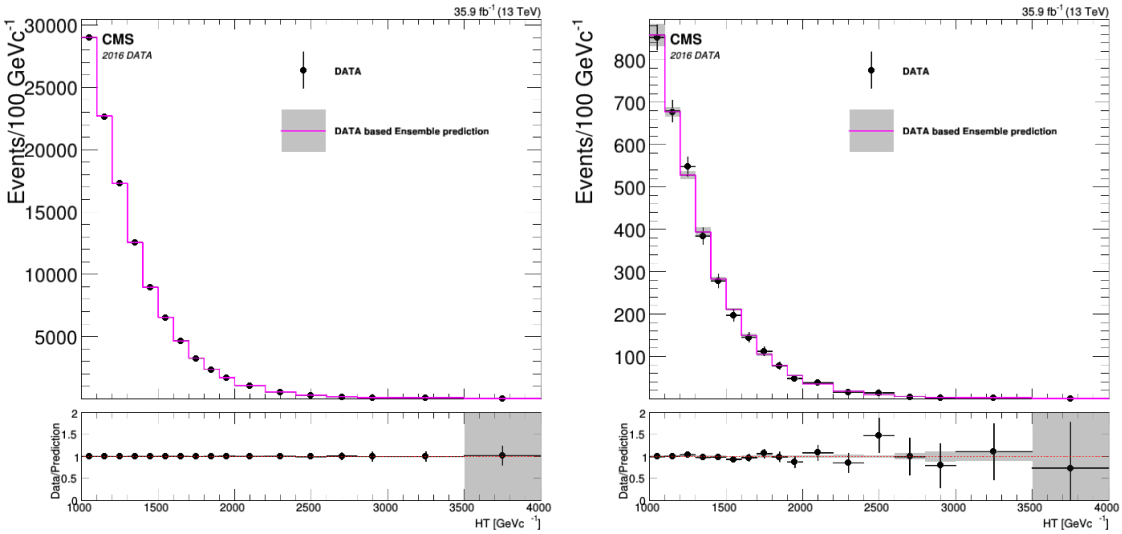

Figure 7.6: Full data H_{T} distribution of events in LT 00 bin (top) and to to the bottom the LT 10 bin for the 4+ jet category

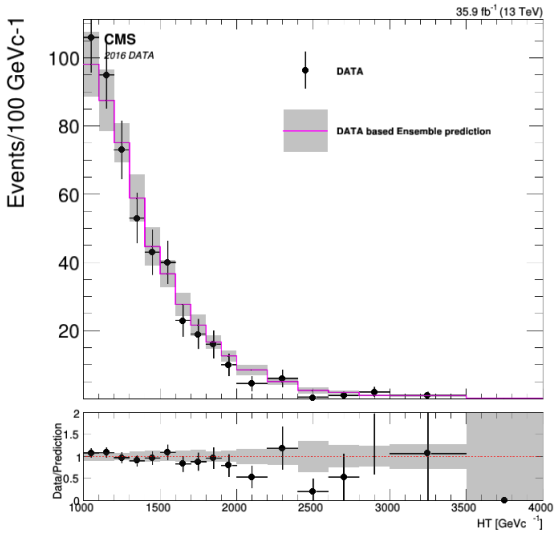

Figure 7.7: Full data H_{T} distribution of events in the LT 11 bin for the 4+ jet category

Chapter 8

Scale Factor for Photon Jets

The interpretation of the results either as upper limits on the cross section of a specific signal model, in the case no excesses are observed, or hypothesis testing observed excesses between different signal models, depends on the ability of the MC simulation to correctly predict how signal events will look like in data. Corrections are derived and applied on MC to ensure agreement between data and simulation. The correction is what is usually called as scale factor. In this chapter we describe the procedure of measuring scale factors for the Loose and Tight photon jet selections we have designed. The results are presented and the application of the scale factor is also explained.

8.1 Tag and Probe method

To measure the scale factor, one needs to identify a process in data that resembles the simulated object or process and use it to calibrate the selections and algorithms developed. In the case of the search presented in this thesis, this would mean to find a SM process that results in merged jets of photons and two gluons, coming from the decay of a boosted object. The closest process that resembles our signal, are top quark decays that merge into a single jet. Specifically top decays to an electron and a b quark [44] with an FSR gluon. The differences between a reconstructed final state photon and an electron, is the presence of a good track for electrons. This final state provides for an EM objects that is embedded in a jet with two other cores to calibrate the Loose and Tight photon jets on.

To select top decays of interest, we use a tag and probe method. This works in cases

where two objects are known to be produced in tandem and it is easy to identify one of the two (the tag) creating a relatively pure sample. The other object, the probe is then used to measure the quantity of interest. We tag top decays to a muon, a b quark and p_{T}^{miss} that recoil against a fat jet which we define as our probe. Top-antitop quark production at the collider yields such events and these selections can create a clean sample limiting other SM processes. The probe jets in this case, are composed of the desired top decay, but also hadronic top decays and initial state radiation jets recoiling off the top-antitop system. Events passing the tag and probe selection also include a limited contribution of other SM processes unrelated to top decays such as QCD jets and W+ jets. The Loose and Tight photon jets selections, require that the photon object in the jet passes the pixel seed veto. We reverse this requirement to actually select electrons and not photons from the top decay.

We use a template fitting method to extract from the selected data probes, the probes corresponding to top to electron decays. The Loose and Tight photon jet selections are then applied on the probes and based on how many of them pass the selections, a data efficiency for each of the selections can be calculated and compared to one measured in simulation, for the extraction of the scale factor. In the next section this procedure is described in detail.

8.2 Template Construction and Fitting procedure

Mass distributions for the processes that contribute in the probe dataset, are modeled using simulation. This includes QCD jets, W+jets, random initial state radiation jets and the top jets. The soft drop mass distributions are constructed then by applying the Tag and Probe selections on the simulation and looking at the probe mass distributions. The templates are then used to find the yield of top probes in the data by fitting the templates to the observed distribution. We do not split the top decays into different templates, but rather keep them as a single component to obtain the inclusive top probes yield from data. The reason for doing this is that the top to electron decays that result sin merged jets are a small fraction to the total top decays and the fitting procedure would not result in a good estimation of

this component. The fraction of the top to electron decays in the probes, is about 4% as MC studies indicate and varies with the probe jet p_{T}. We use this number to calculate the total observed top to electron decays in the data probes , by taking the yield of the inclusive top probes from the fit and then multiplying it with the MC measured fraction. This fit as well as the following ones, are performed in p_{T} bins, in which we split the observed probe jets. Thus we obtain at the end three scale factor measurements, one for each of the three p_{T} bins we have specified. The three probe jet p_{T} bins are 200-280 GeV, 280-380 GeV and the final 380+ GeV.

The data probes are then required to pass the Loose Photon jet selection, with the pixel seed veto flipped. Once more we obtain a distribution of probes that pass the selections. This distribution is mostly composed of top decays, a small component of W+jets and further an even smaller QCD component. The data probes are fitted here with MC derived templates. The templates are representative of the following processes, top to electron decays, top to non electron decays, W+jets and QCD jets. The top contribution is split here since it becomes a considerable component and the fitting procedure can easily disentangle correctly how many data probes are actual top to electron decays. This yield is then divided by the yield for top to electron decays from the first fitting procedure of all the probes and not only the ones passing the Loose photon jets selection. The ratio is the efficiency of the Loose photon jet selection on top to electron decays. This can be then divided by the MC measured efficiency for top to electron decays to pass the Loose photon jet selection and obtain the scale factor.

To calculate the Tight photon jets efficiency, an extra template fit is required. First we take the data probes that passed the Loose photon jets selection with the flipped pixel veto requirement and proceed with applying the Tight photon jets selections on them. The data probes that fulfill this selection, are composed mostly of top decays to electrons, some top hadronic decays and an even smalled QCD jets component. MC templates are once more used to fit the probe jet mass distributions to obtain the yield of top to electron decays

probes in data. This yield is then divided by the result of the second template fitting procedure on Loose photon jet probes and the result is the efficiency for top to electron decays to be labeled as Tight photon jets. As with the Loose photon jets result, a ratio of the data measured efficiency with the MC efficiency for top to electron decays is the scale factor that we need.

8.3 Alternative method

The template fitting procedure can yield results with high uncertainties, because of the fact that the templates we use for the fits are correlated. The hadronic top and top to electron decays templates are highly correlated since they are derived from the same MC simulation and a matching procedure is used to label the probe jets as the one or the other using MC generated particle information. Another correlation between templates, is that the template for Tight photon jets are constructed by applying the Tight photon jet selection on the Loose photon jet probes - that make up the template used for the Loose photon jet fit. Beyond having correlated templates, performing fits with all the components being let free to fluctuate, where the distribution shape is the main driver of the fit result, returns large uncertainties. As a remedy to this, we fix the yield of top hadronic decays using a data driven method, and then perform the fit.

The way to constrain the hadronic top decays, is by realizing that if we apply the nominal Loose and Tight photon jet selections on data, the top component that survives, is actually the hadronic top decays, since the pixel veto kills all the top to electron decays. Thus we can apply this selection on the probes distribution in data, perform a fit and obtain the yield of top hadronic decay passing our nominal Loose and Tight photon jet selections. The idea is that if we then have an efficiency for the top hadronic decays to fail the pixel veto requirement, we could use it to multiply it with this yield and obtain how many hadronic top decays pass the modified Loose/Tight photon jet selections. Luckily the pixel seed selection is well modeled in data with the CMS experts measuring the difference between

the behavior of this selection in data and MC to be in order of a few %. We then measure in simulation how many top hadronic decays pass the nominal selections and how many fail it and the ratio is used to estimate the top hadronic decays in the Loose and Tight photon jets probe in data with the pixel veto reversed. The result of this procedure is a scale factor measurement with reduced uncertainties that we use as our final scale factor result. The results are presented in fig. 8.1 and table 8.1. The fit error corresponds to the total error from the template fitting procedure. The final scale factor is taken by considering the three p_{T} scale factor measurements and fitting them with a constant term to take into account their error and a single value for the total p_{T} range.

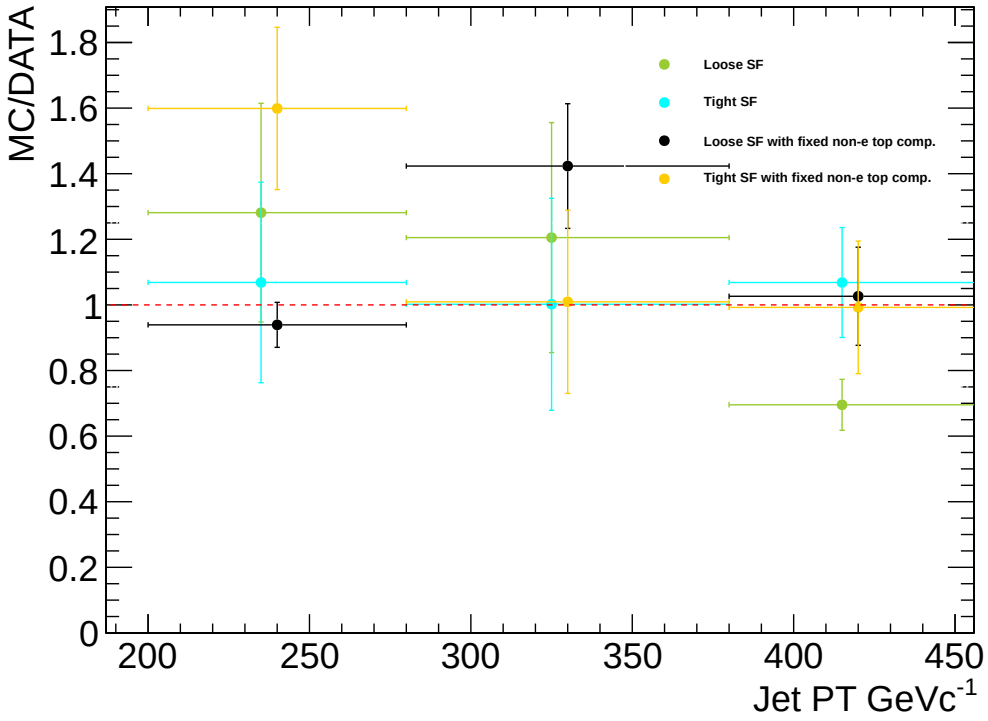

Figure 8.1: These two method for measuring the scale factor with and without fixed top hadronic decays are shown. The uncertainty in the method with fixed top hadronic decays, is considerably smaller and those results are used as the signal efficiency correction

	Loose Rate	Tight Rate
Scale factor	1.1 ± 0.1 (fit)	1.2 ± 0.1(fit)

Table 8.1: The scale factor results used for the method calibration.

8.4 Hadronizer uncertainty

One can not tag in data top decays to electrons, b quarks and final state gluons radiated of the b quark. Our result is depended on the top to electron decays with the extra radiation. For this reason, we apply an uncertainty on the scale factor based on the modeling efficiency of MC of this extra final state radiation. To test this and quantify the uncertainty from this effect, we use top decays hadronized with two different packages. The first which was used for the scale factor result, is done with Pythia8 while the second is done using the Herwig package. Signal events are also hadronized with Herwig. The Loose photon jet efficiencies for top decays are then measured from the two hadronizers and a ratio is calculated. The signal efficiency to pass the Loose photon jet selection, is also compared between the two hadronizers, using our nominal signal MC sample, hadronized with Pythia and the Herwig hadronized signal mass points. A ratio of the two ratios is then calculated as a function of the probe jet or signal jet p_{T}. The deviation from unity is used as an extra uncertainty on the scale factor. Since the top probe jets distribution is limited in high p_{T}, an extrapolation of the ratio is performed from the low p_{T} values. This results in an extra 20% uncertainty for the low p_{T} probe jets on the scale factor, raging up to about 70% in the high p_{T} range - close to the 2 TeV values. This uncertainty is then added to the fit uncertainty as the final scale factor uncertainty. Luckily signal events, even though they have high energetic jets, the bulk does not get affected by this large uncertainty, with most of the events obtaining a 20-30% extra uncertainty on the scale factor.

8.5 Scale factor application

Once we have the scale factor and it's uncertainties, then it is possible to scale the signal expectation shapes to correct for the data-to-simulation difference. The scale factor we have measured is a per jet quantity, that needs to be translated to an event weight, to correct the H_{T} distributions of the signal region. We use the scale factor as a pseudo probability.

Every event has an X number of jets passing the Loose and Tight Photon jets selections. The weight for this particular event is the calculated as $(scalefactor)^X$. The uncertainties of the scale factor, are propagated to the H_{T} shapes as variation in the scale factor we use. The scale factor is shifted up and down based on the fit and hadronizer uncertainty and an new weight is constructed for variation shifted up and a variation shifted down. The weighted distributions with thew varied scale factors, are introduced to the limit setter as nuisance parameters. The total scale factor uncertainty on signal events is found to be between 30% to 50%, being the largest systematic uncertainty in this search.

By measuring the scale factor and correcting the simulation signal efficiencies, we can trust and quantify the uncertainty on either the upper limits we will set or the hypothesis testing we will perform between different models.

Chapter 9

Systematic Uncertainties

Interpreting the results with a statistical analysis, either as consistency testing between observed data and background expectations, or as upper limits on cross-section production, or even as hypothesis testing between different signal models, needs to account for sources of uncertainties and their effects on the final results. In the search presented in this thesis, a number of sources of systematic uncertainties are considered and incorporated in the statistical treatment of the observed data. All the uncertainties and their effect are listed in table 9.1. Uncertainties, can either be introduced as scale uncertainties, that is an uncertainty on the total yield of the signal events in the signal region, or they can be a shape uncertainty, that is they indicate how a distribution shape can vary.

For the signal expectation shapes, we test effects related to the simulation of the signal, detector calibration effects as also other sources of uncertainty. The first source of uncertainty studied for the signal process is the uncertainty on the photon jets scale factor. The uncertainty from the fitting procedure that measures the scale factor, and the additional hadronizer uncertainty, are introduced as shape uncertainties to the statistical analysis. This source of uncertainty is the largest in this search.

Signal shapes are normalized to the integrated luminosity, to predict how many signal events will populate the signal region. The integrated luminosity is measured in data and there is an associated uncertainty.

The signal simulation, was performed with the FastSIM package that emulates the detector performance and reconstruction of objects in faster and an approximate way. The reason that this package was used is that the needed phase space for the signal masses

was extensive and thus it was computationally demanding to simulate the signal points using the FullSIM package that simulates the detector response in more detail. The signal simulation assumes a specific pile-up profile. Signal events then are weighted to have the simulation pile-up profile, matching the data observed pile-up profile. An uncertainty related to pile-up reweighting, is incorporated in the statistical analysis, related to the error on the pile-up profile measurement in data.

Another source of uncertainty for the signal simulation, comes from the choice of the parton distribution functions for the process generator, since it changes the energy of the interacting partons that generate the pair of gluinos in the detector. A study with different PDFs is performed to estimate this uncertainty.

The background estimation for the signal region is also associated with an uncertainty calculated by the ensemble method.The reconstructed jets, are calibrated to correct for detector effects that could lead to mis-measurements of the true energies. Simulated jets, undergo a smearing procedure, that is their reconstructed energy in simulation is smeared to match the data reconstructed energy resolution. Both methods have uncertainties associated with them that are considered in the statistical analysis.

If there is recoiling initial state radiation (ISR) on the generated gluino gluino pair, this can result in a Lorentz boost for the gluinos. This in turn can affect the distribution of the final objects in the detector. We performed a study to see how this extra ISR affects the final signal H_{T} distributions in the signal region. The ISR effect was found to be negligible. The trigger efficiency measurement, is also found to be a negligible and is not incorporated that statistical analysis.

Source	Magnitude
simulation-to-data signal efficiency correction $*$	$30 - 50\%$
background estimation $*$	10%
jet energy resolution $*$	$< 10\%$
jet energy scale corrections $*$	$< 10\%$
pileup re-weighting $*$	$< 5\%$
integrated luminosity	2.5%
detector FullSIM - FastSIM	$1 - 2\%$
PDF choice uncertainty	1%

Table 9.1: Systematic uncertainties considered. The asterisk ($*$) denotes uncertainties that are treated as shape uncertainties, while the rest as uncertainties on the normalization. The magnitude of each uncertainty refers to the effect on signal event efficiencies.

Chapter 10

Results and Conclusion

A *blinded* search, is an analysis, that has been designed and optimized, with out considering how data look in the search window during the design process. The signal expected shapes from simulation with the simulated background processes and the characteristics of the data from the background dominated region are taken into account, to design the analysis while keeping the signal region blinded. The idea behind this tactic is to avoid analysts being biased by the shape of data designing cuts and selections to either enhance or remove perceived features from the observed data. Once the analysis features have been defined optimized and tested, with the background estimation method to have been proven to work and further tests that the analysis would indeed be sensitive to new physics existing in the signal region, the blindfold is lifted. The data distributions from the signal region are revealed and the results are examined to determine if evidence for new physics are present. This process is known as *unblinding* by particle physicists.

The unblinded data distributions from the signal region of events with three AK8 jets and events with 4+ AK8 jets are presented in figures 10.1 and 10.2, respectively. The H_{T} distributions of data seem to be in agreement with the background predictions. The total event yields in each of the 6 total signal region bins, are presented in table 10.1. In the 4+ jet category the LT 22 bin, is predicted to have one event based on the background expectation, while zero are observed. No significant excess is observed. In fig. **??** the event display of an event from the 3jet category from the LT-22 signal region with an H_{T} of 1.2 TeV is presented.

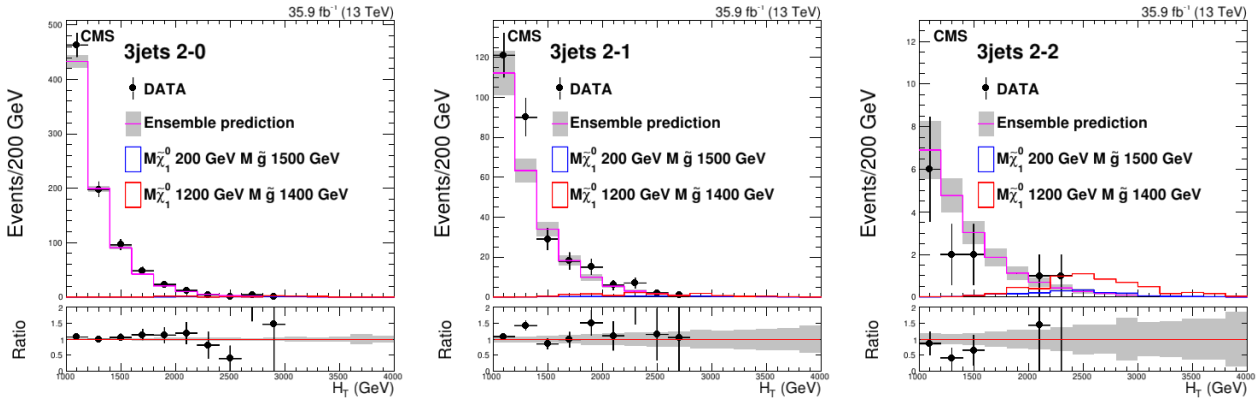

Figure 10.1: Data observed H_{T} distributions in the signal search window for the 3 AK8 jet category.

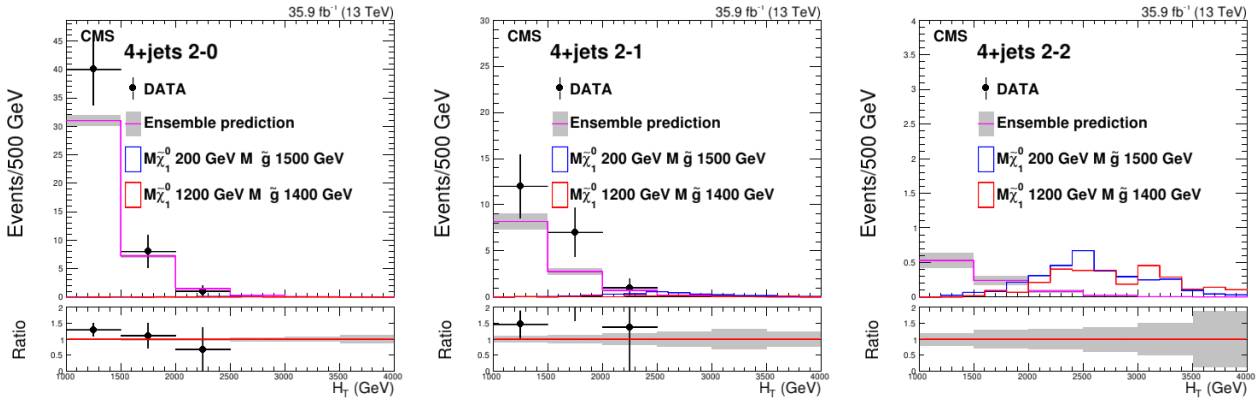

Figure 10.2: Data observed H_{T} distributions in the signal search window for the 4+ AK8 jet category. In the LT 22 category, one event is predicted by the background estimation method, while zero are observed.

Bin	Expected	Observed
	3 jets category	
LT20	805.6 +22.2 −23.8	850
LT21	249.8 +27.9 −25.0	289
LT22	19.5 +4.1 −3.5	12
	4+ jets category	
LT20	40.1 +1.3 −1.2	49
LT21	11.9 +1.4 −1.4	20
LT22	0.9 +0.2 −0.2	0

Table 10.1: Observed event yields in the different event categories of the Signal Region. The expected event yields based on the ensemble method prediction are compared to the observed event yields for each of the categories. The top section shows results for the 3 jets events category and at the bottom the 4+ jets category is presented.

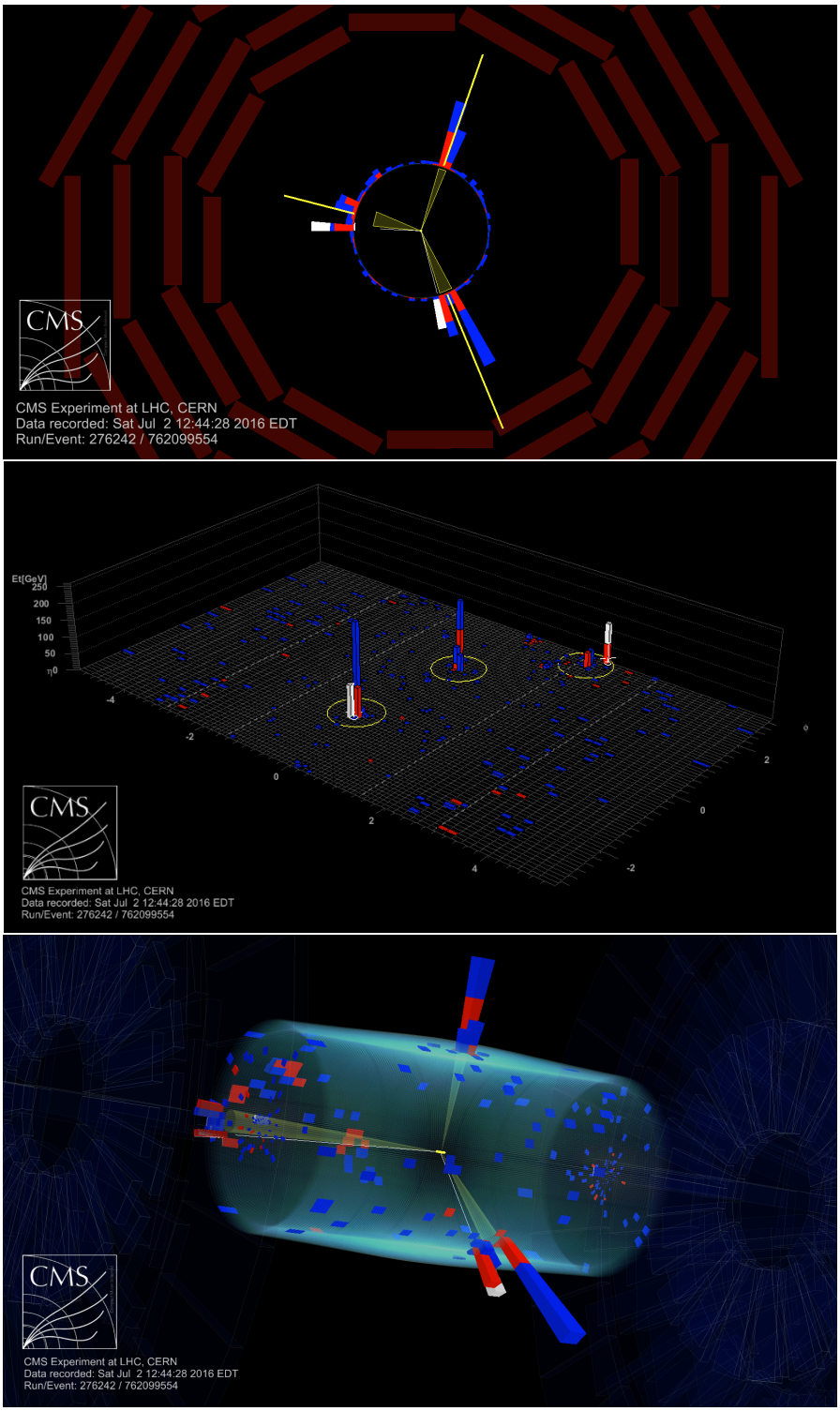

Figure 10.3: An event display for an event from the 3 AK8 jet Signal Region. The red color represents energy depositions in the ECAL. Blue color represents energy depositions in the HCAL while the white present the calorimeters associated with photons. The yellow triangles present the reconstructed AK8 jets in the event with $p_{\mathrm{T}} > 200$ GeV. Two projections of the detector are presented - an $\rho - \phi$ projection and the $\phi - \eta$ view. The third is a 3-D representation of the observed event. The event has an H_{T} of 1.12 TeV.

Following the result presentation, in the next section the statistical analysis and interpretation of the observed data is discussed.

10.1 Statistical Analysis and Interpretation

We use the observed data H_{T} distribution, to set upper limits on the cross section of gluino pair production. A Bayesian method is used to calculate limits on the signal strength that are then transformed to cross -section limits [50]. Limits for every signal mass point are calculated at a 95% confidence level and the mass space exclusion space is constructed. Signal mass points are excluded if the upper limit of the cross section calculated, is less than the theoretical predicted value.

The results of the excludes phase-space are presented in fig. 10.4 for the three AK8 and 4+ AK8 jets categories separately while a joint analysis is performed and the results are shown in fig. 10.5 . Two effects shape the excluded mass phase space. The first is the gain in acceptance in the boosted regime from the usage of the jet substructure techniques and the photon subjet fraction. That extends the excluded phase space for low mass neutralinos (below 300 GeV) up to 1.7 TeV for gluino masses. For neutralino masses above 300 GeV, the excluded phase space retreats to lower gluino masses, since the mass gap between the two particles becomes smaller and the neutralino stop being as boosted as below 300 GeV. For neutralino masses below 150 GeV the final states become over collimated that even the substructure techniques we use can not discern and identify them and thus the excluded phase space retreats again to lower gluino mass values.

The second effect into play that affects the exclusion space, is accidental substructure, mainly coming fro the 3 AK8 jets category. Final states coming from the two gluinos start mixing together and merging into single jets, ie a photon from the neutralino decay of one gluino and the singlet decays coming from the other neutralino in the event. By looking at events with only 3 AK8 events, means that signal events that pass this selection, will have final states collimated together, which benefits the large neutralino mass signal points,

(1 TeV) and thus the extended exclusion in the large neutralino large gluino masses.

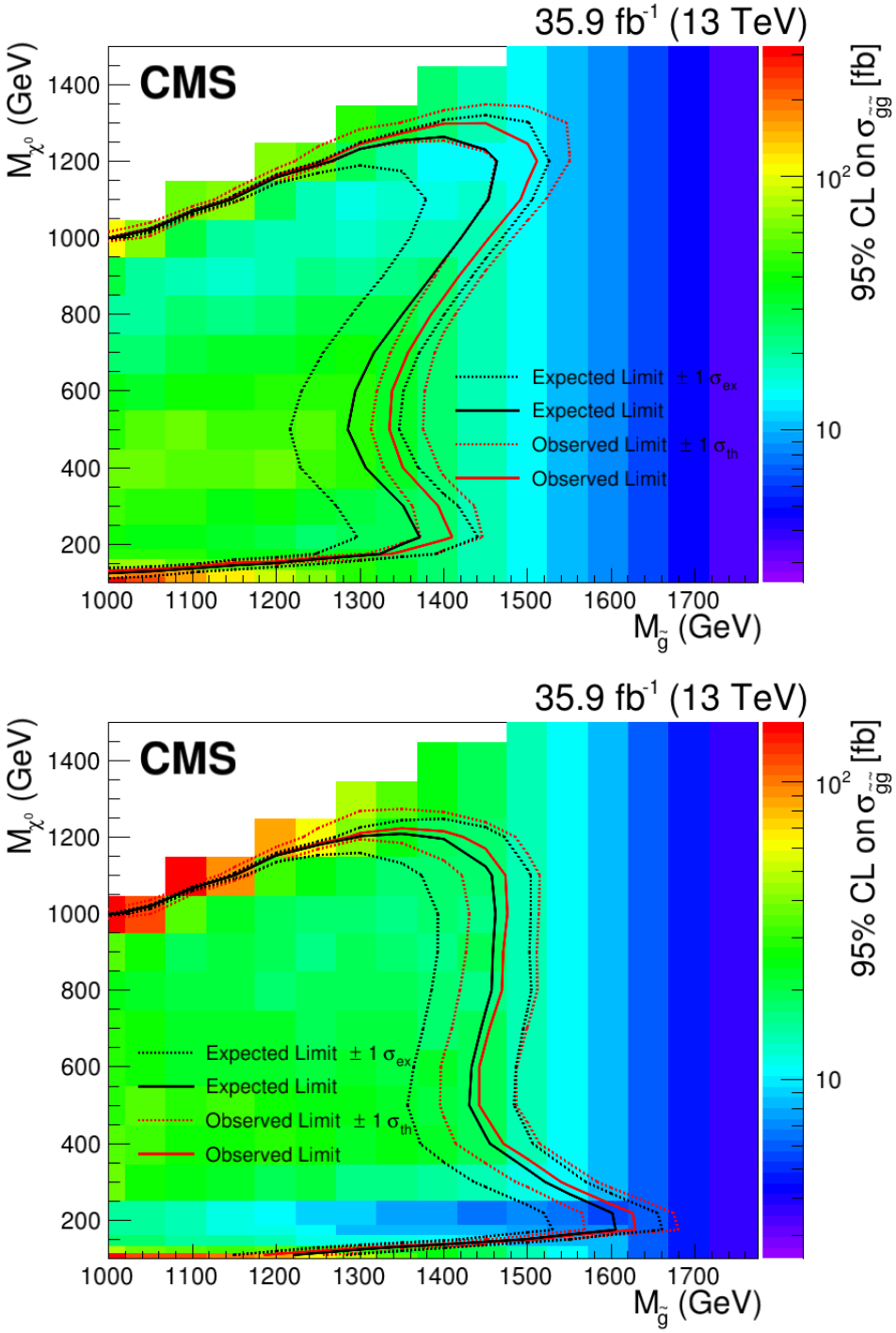

Figure 10.4: Excluded phase-space for the 3 AK8 jet (top) and 4+ AK8 jet (bottom) event categories. The expected excluded space is presented too with the black line while the red line denotes the exclusion based on the observed data. Expected and Observed limits are within 1 σ to each other.

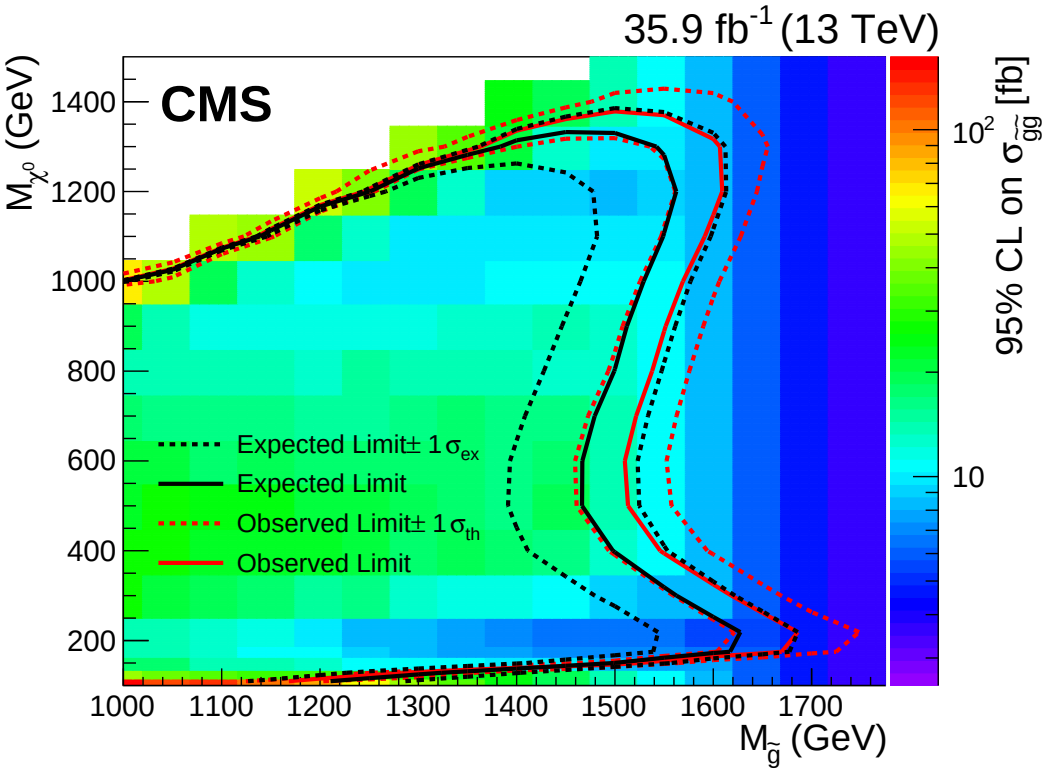

Figure 10.5: Excluded mass space based on the joint statistical analysis of NJ3 and NJ4+ events. The excluded mass-space is enclosed by the exclusion lines. The observed limits are represented with a red line are within the 1σ to the expected limits (black line). The error on the observed limit corresponds to the theoretical uncertainties on the cross-section for the gluino pair production.

10.2 Conclusion

A search for new physics for events with collimated photons and gluons into jets, is presented in this thesis. The search is performed on data collected by the CMS experiment during the 2016 Run 2 of the LHC and correspond to an integrated luminosity of 35.9 fb^{-1}. The data are proton-proton collision events, at a center-of-mass energy of 13 TeV. For the first time the search probes the multijet QCD background, using photons in jets as a handle to suppress the standard model background and uncover new physics. Using photons in jets is a new tool that can have many applications, from discovering new physics, to measuring SM processes that are tough to observe because of high boost.

No statistically significant evidence for new physics is found. The results are interpreted as upper limits on the cross-section for the pair production of gluinos, excluding a phase-space of possible gluino masses up to 1.7 TeV. The boosted regime previously unexplored by searches using isolated final states, is investigated. Even though new phenomena have not been uncovered with this search, a road map for analyses that attempt to probe the large mutlijet standard model background is set. New physics can still be hiding in the record amount data collected by the LHC, and until we probe every single corner of them, the final word about whether nature is fine-tuned or not has not been said yet.

Appendix A

Datasets and Simulation Samples

This section describes the samples used from RUN 2 and the Simulation samples we used for this analysis. Table A.1 describes the H_{T} QCD samples used. The Signal samples along with $\bar{t}$ and WJets are described in table A.1. The JetHT primary datasets used for the search are listed in A.5 while A.3 describes the datasets used for the measurement of the scale factor of our algorithm. The cross-sections for the different signal mass points (fixed m $\widetilde{\mathrm{g}}$) used are taken from [12].

Sample description	CMS DAS Name	Events	$\sigma \times B$ [pb]
QCDH_{T} 0.3-0.5 TeV	/QCD_HT300to500_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1/	17035891	351300.0±1682
QCDH_{T} 0.3-0.5 TeVext	/QCD_HT300to500_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6_ext1-v1/	37483250	351300.0±1682
QCDH_{T} 0.5-0.7 TeV	/QCD_HT500to700_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1	18689717	31630.0±937
QCDH_{T} 0.5-0.7TeVext	/QCD_HT500to700_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6_ext1-v2	42790025	31630.0±937
QCDH_{T} 0.7-1TeV	/QCD_HT700to1000_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1	15629253	6802.0±43
QCDH_{T} 0.7-1 TeV ext	/QCD_HT700to1000_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6_ext1-v1	29783527	6802.0±43
QCDH_{T} 1 -1.5 TeV	/QCD_HT1000to1500_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2_ asymptotic_2016_TrancheIV_v6-v1	4756101	1206.0 ± 6.5
QCDH_{T} 1 -1.5 TeV ext	/QCD_HT1000to1500_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2_ asymptotic_2016_TrancheIV_v6_ext1-v1	10360193	1206.0±6.5
QCDH_{T} 1.5 -2 TeV	/QCD_HT1500to2000_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2_ asymptotic_2016_TrancheIV_v6-v1	3970819	120.4±0.64
QCDH_{T} 1.5 -2 TeV ext	/QCD_HT1500to2000_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6_ext1-v1	7855883	120.4±0.64
QCDH_{T} 2- inf TeV	/QCD_HT2000toInf_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1	1991645	25.25±0.15
QCDH_{T} 2-inf TeV ext	/QCD_HT2000toInf_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6_ext1-v1	4047360	25.25±0.15

Table A.1: QCD simulated samples utilized and their cross sections.

Sample description	CMS DAS Name	Events	$\sigma\times$ B [pb]
$t\bar{t}$	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8 /RunIISummer16MiniAODv2-PUMoriond17_80X _mcRun2_asymptotic_2016_TrancheIV_v6-v1/	77081156	831.76
Signal gluino	/SMS-T7WgStealth_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUSummer16Fast_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1/	45973666	$\star$
WJtoQQ	/WJetsToQQ_HT-600ToInf_TuneCUETP8M1_13TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1/	1026587	1449 $\pm$ 2
WJtoLNu	/WJetsToLNu_TuneCUETP8M1$_1$3TeV -madgraphMLM-pythia8/RunIISummer16 MiniAODv2-PUMoriond17_80X_mcRun2 _asymptotic_2016_TrancheIV_v6-v1/	29705748	50690 $\pm$ 3891

Table A.2: Signal simulation samples, $t\bar{t}$ and W+ jets utilized. The cross-sections used for the different $\tilde{g}$ mass points, are taken from [12] .

Dataset description	CMS DAS Name	Int Lumi [pb^{-1}]
RUN B	/SingleMuon/Run2016B-03Feb2017_ver2-v2	5780.499
RUN C	/SingleMuon/Run2016C-03Feb2017-v1	2573.399
RUN D	/SingleMuon/Run2016D-03Feb2017-v1	4248.384
RUN E	/SingleMuon/Run2016E-03Feb2017-v1	4008.376
RUN F	/SingleMuon/Run2016F-03Feb2017-v1	3101.618
RUN G	/SingleMuon/Run2016G-03Feb2017-v1	7540.488
RUN H	/SingleMuon/Run2016H-03Feb2017_ver2-v1	8390.540
RUN H	/SingleMuon/Run2016H-03Feb2017_ver3-v1	215.149

Table A.3: Single Muon Primary Datasets utilized for the scale factor extraction.

Dataset description	CMS DAS Name	Int Lumi [pb^{-1}]
RUN B	/SingleElectron/Run2016B-03Feb2017_ver2-v2	5780.499
RUN C	/SingleElectron/Run2016C-03Feb2017-v1	2573.399
RUN D	/SingleElectron/Run2016D-03Feb2017-v1	4248.384
RUN E	/SingleElectron/Run2016E-03Feb2017-v1	4008.376
RUN F	/SingleElectron/Run2016F-03Feb2017-v1	3101.618
RUN G	/SingleElectron/Run2016G-03Feb2017-v1	7540.488
RUN H	/SingleElectron/Run2016H-03Feb2017_ver2-v1	8390.540
RUN H	/SingleElectron/Run2016H-03Feb2017_ver3-v1	215.149

Table A.4: Single Electron Primary Datasets utilized for the trigger effieciencies studies.

Dataset description	CMS DAS Name	Int Lumi [pb^{-1}]
RUN B	/JetHT/Run2016B-03Feb2017_ver2-v2/	5784.727
RUN C	/JetHT/Run2016C-03Feb2017-v1/	2573.399
RUN D	/JetHT/Run2016D-03Feb2017-v1/	4248.384
RUN E	/JetHT/Run2016E-03Feb2017-v1/	4008.376
RUN F	/JetHT/Run2016F-03Feb2017-v1/	3101.618
RUN G	/JetHT/Run2016G-03Feb2017-v1/	7540.488
RUN H	/JetHT/Run2016H-03Feb2017_ver2-v1/	8390.540
RUN H	/JetHT/Run2016H-03Feb2017_ver3-v1/	215.149

Table A.5: JetDatasets utilized for the main search.

Appendix B

Trigger Studies

In this appendix the efficiency curves for the HLT_PFHT_900 trigger for each Run separately are presented in fig.B.1 to B.4. The Trigger efficiency for the triple trigger OR combination that we use to recover the losses of RunH is plotted versus the leading AK4 pt jet in the event in fig B.5.

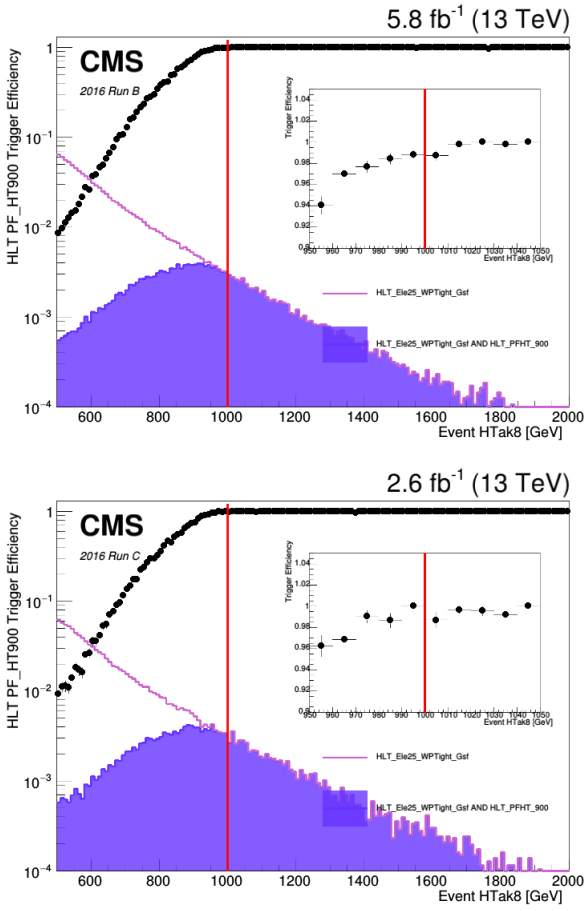

Figure B.1: The Trigger efficiency for RUNs B and C

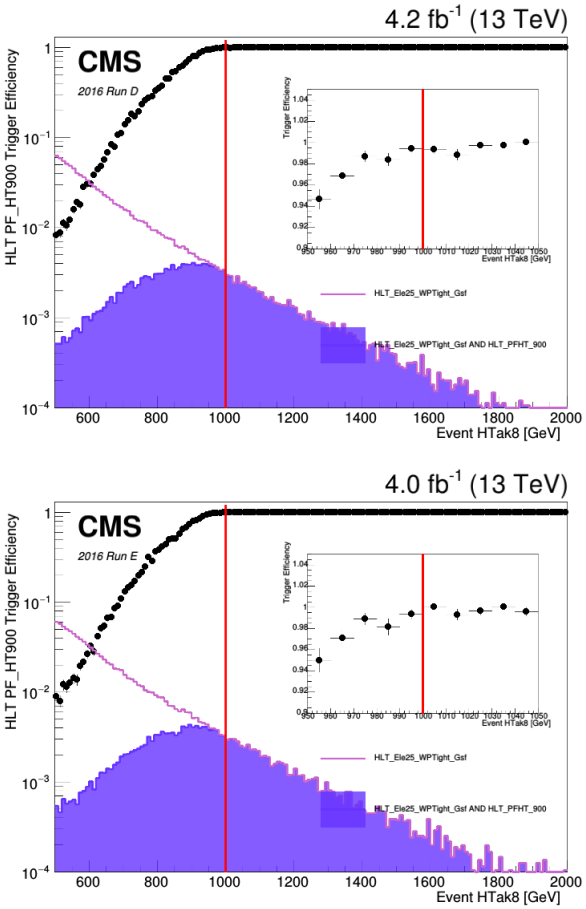

Figure B.2: The Trigger efficiency for RUNs D and E

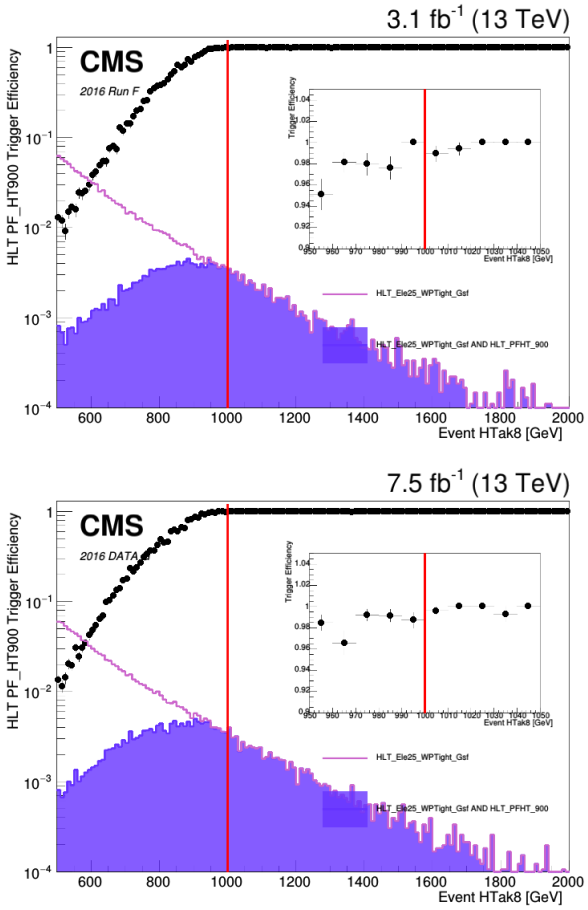

Figure B.3: The Trigger efficiency for RUNs F and G

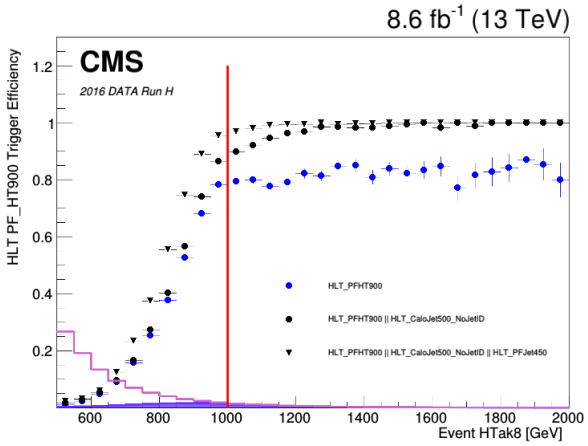

Figure B.4: The Trigger efficiency for RUN H

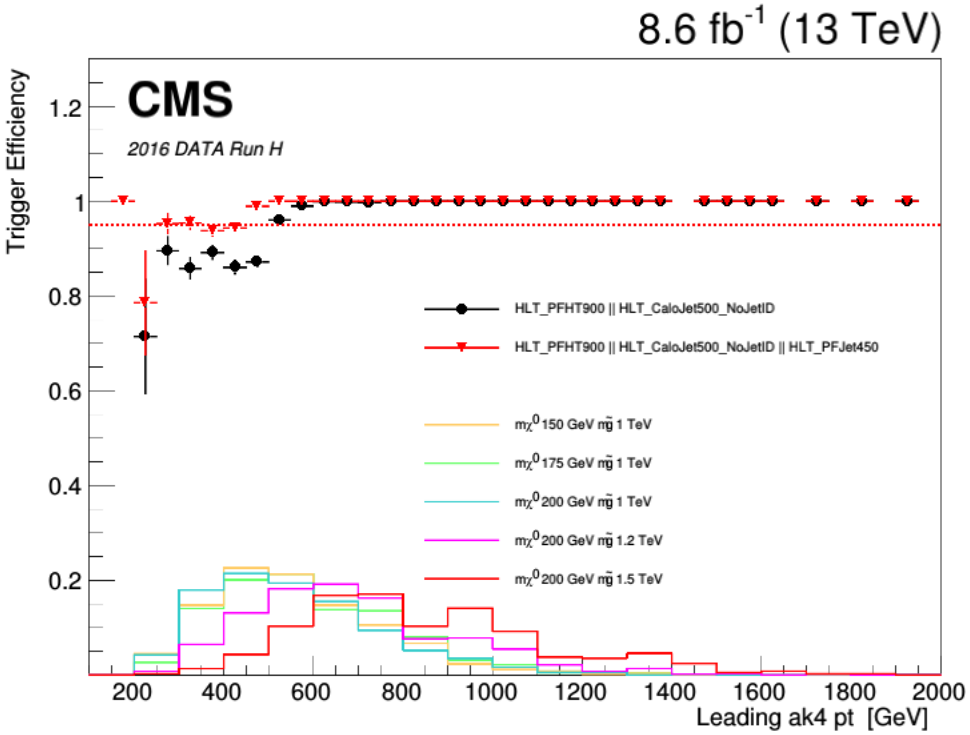

Figure B.5: The Trigger efficiency vs the leading ak4 jet pt of each event with an $H_{\mathrm{T}} > 1$ TeV. The different trigger combinations are presented and the dashed red line corresponds to a 95% efficiency line. A weight based on the H_T effect of the inefficiency is applied on the expected signal shapes for the RUNH fraction of the total events.

Bibliography

[1] Morad Aaboud et al. Measurement of the W^+W^- production cross section in pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with the ATLAS experiment. *Phys. Lett.*, B773:354–374, 2017.

[2] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[3] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. Perrotta. The fast simulation of the CMS detector at LHC. *J. Phys. Conf. Ser.*, 331:032049, 2011.

[4] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[5] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[6] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[7] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473–500, 2008.

[8] M. Bahr et al. Herwig++ physics and manual. *Eur. Phys. J. C*, 58:639–707, 2008.

[9] R. Barbier, C. BÃľrat, M. BesanÃğon, M. Chemtob, A. Deandrea, E. Dudas, P. Fayet, S. Lavignac, G. Moreau, E. Perez, and Y. Sirois. R-parity-violating supersymmetry. *Physics Reports*, 420:1, 2005.

[10] David Barney. CMS Detector Slice. CMS Collection., Jan 2016.

[11] G. L. Bayatian et al. CMS Physics. 2006.

[12] Christoph Borschensky, Michael KrÃđmer, Anna Kulesza, Michelangelo Mangano, Sanjay Padhi, Tilman Plehn, and Xavier Portell. Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13$, 14, 33 and 100 TeV. *Eur. Phys. J. C*, 74(12):3174, 2014.

[13] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[14] S. Catani, Yuri L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant k_{T} clustering algorithms for hadron hadron collisions. *Nucl. Phys. B*, 406:187, 1993.

[15] Serguei Chatrchyan et al. Search for supersymmetry in events with photons and low missing transverse energy in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 719:42, 2013.

[16] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[17] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[18] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[19] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[20] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[21] Wikimedia Commons. File:elementary particle interactions.svg — wikimedia commons, the free media repository, 2016.

[22] Wikimedia Commons. File:standard model of elementary particles.svg — wikimedia commons, the free media repository, 2019.

[23] Valeriane Duvivier. Cross section of LHC dipole.. Dipole LHC: Coupe transversale. AC Collection. Legacy of AC. Pictures from 1992 to 2002., 2001.

[24] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[25] JiJi Fan, Rebecca Krall, David Pinner, Matthew Reece, and Joshua T. Ruderman. Stealth supersymmetry simplified. *JHEP*, 07:016, 2016.

[26] JiJi Fan, Matthew Reece, and Joshua T. Ruderman. Stealth supersymmetry. *JHEP*, 11:012, 2011.

[27] JiJi Fan, Matthew Reece, and Joshua T. Ruderman. A stealth supersymmetry sampler. *JHEP*, 7:196, 2012.

[28] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[29] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007.

[30] Andrea Giammanco. The Fast Simulation of the CMS Experiment. *J. Phys. Conf. Ser.*, 513:022012, 2014.

[31] Stefan Gieseke, Christian Rohr, and Andrzej Siodmok. Colour reconnections in Herwig++. *Eur. Phys. J. C*, 72:2225, 2012.

[32] Vardan Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at âĹŽs = 8 TeV. *JINST*, 10(06):P06005, 2015.

[33] Vardan Khachatryan et al. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV. *JINST*, 10(08):P08010, 2015.

[34] Vardan Khachatryan et al. Search for stealth supersymmetry in events with jets, either photons or leptons, and low missing transverse momentum in pp collisions at 8 TeV. *Phys. Lett. B*, 743:503, 2015.

[35] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76:155, 2016.

[36] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12:P02014, 2017.

[37] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12(01):P01020, 2017.

[38] Christiane LefÃĺvre. The CERN accelerator complex. Complexe des accÃľlÃŀrateurs du CERN. Dec 2008.

[39] Stephen P. Martin. A Supersymmetry primer. page 1, 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[40] Stephen Myers. *The LEP Collider, from design to approval and commissioning.* John Adams' Lecture. CERN, Geneva, 1991. Delivered at CERN, 26 Nov 1990.

[41] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[42] NNPDF Collaboration. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244, 2013.

[43] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12:P10003, 2017.

[44] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13:P05011, 2018.

[45] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018.

[46] TorbjÃűrn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015.

[47] W.J. Stirling. private communication.

[48] Matthew J. Strassler. Possible effects of a hidden valley on supersymmetric phenomenology. 2006.

[49] Matthew J. Strassler and Kathryn M. Zurek. Echoes of a hidden valley at hadron colliders. *Phys. Lett. B*, 651:374, 2007.

[50] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev. D*, 98:030001, 2018.

[51] J. Thaler and K. Van Tilburg. Identifying boosted objects with N-subjettiness. *JHEP*, 3:15, 2011.

[52] Robert Rathbun Wilson. The Tevatron. *Phys. Today*, 30N10:23–30, 1977.